The Effectiveness of using Static Features in Identifying Scam Genres

AMBER C. STABEK

This thesis is submitted in total fulfilment of the requirements for the degree of

Master by Research of Mathematical Sciences

Graduate School of Information Technology

and

Mathematical Sciences

University of Ballarat

University Drive, Mt Helen

Submitted February 15, 2010

Abstract

Variation in scam classification is regularly identified as a primary cause of discrepancy in victim report data resulting in unsuccessful scam identification and insufficient rates of interception by law enforcement, which results in the low prosecution rate of scammers. The result of such discrepancies lead to complex concerns, such as the under reporting of scam incidence, and reduced rates of successful follow up by investigative and enforcement agencies consequential to difficulties in making correct referrals. Without a shared and common lexicon of scam labels and descriptions, communication between investigative agencies and cross-border cooperation is obstructed. With no compatible comprehension of the scam lexicon, timely progression in scam-case management leading to the identification, tracking and interception of scammer communications cannot be realised. Ambiguities leading to interpretational impedances are aiding scammers by enabling their scams in cross-jurisdictional and multi-national platforms. If the wide variety of known scam types could be condensed to recognisable and traceable instances, the business models that scammers use could be identified and future scamming events predicted, monitored, and interrupted.

Following a mixed methodology, this research aims to address some of these concerns. This is achieved by clustering scam descriptions and partitioning them into scam types, called scam genres. The result of which reveals homogeneous groups of scam cases and allows for the assessment of the effectiveness of using static features in identifying scam types. Second to this, identification of the most suitable model for reducing scam cases into the fewest number of clusters with the least number of scam cases within in each cluster at an accuracy level of at least 95% is achieved.

Through the use of hierarchical clustering, this research grouped publically available scams into homogeneous clusters of scam genres. Two-hundred and seventy-seven scams from 38 separate categories of scam classification were condensed into as few as 7-clusters of scam genre. Following a mixed methodological, grounded theoretical approach and using discriminant function analysis, 82 static features were derived from the 277 scam descriptions analysed. Of the 82 static features derived, it was concluded that only 68 significantly predicted scam type and explained 95% of the total variation found in scam case assignment. The most significant static features determined to be crucial to any scamming campaign and useful in identifying the type of scam genre a scam case belongs to were; what the scam offered, the role of the victim, the goal of the scammer and the method of scam introduction.

The results of this research provide empirical evidence of the inconsistent use of definitions across jurisdictions in scam descriptions, and will contribute to the development of a uniform lexicon of scamming terminology as well as become foundational to further research on the impact of scams for law enforcement, the public and private sector, the community and the individual.

Statement of Authorship

Except where explicit reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis by which I have qualified for or been awarded another degree or diploma. No other person's work has been relied upon or used without due acknowledgement in the main text and bibliography of this thesis.

Signed: ______________________________	Signed: ______________________________
Dated: _______________________________	Dated: _______________________________
Amber C. Stabek	Dr. Paul Watters
Candidate	Principal Supervisor

Acknowledgements

The completion of this research was made possible by the wonderful staff of the University of Ballarat, those within the Research and Graduate Studies Office, in particular Diane Clingin and Elanor Mahon, and the Graduate School of Information Technology and Mathematical Sciences. The support offered by the Internet Commerce Security Laboratory team has been invaluable to me and I would like to thank in particular Robert Layton, Mamoun Alazab, and Maxine Kingston for their patience and assistance throughout the last 18 months. I must also thank my Principal Supervisor, Dr. Paul Watters for his unconditional support, respect and belief in me.

This research has been supported by IBM, Westpac, the Victorian Government and the Australian Federal Police, the shared communications between these sectors has been invaluable and has offered insight where blindness would have ensued. A special thank you to Simon Brown of Westpac as well as Detective Superintendant Brian Hay, Operations Commander Fraud and Corporate Crime Group State Crime Operations Command Queensland Police Service, whose friendship and ongoing commitment to victims of crime has both assisted and inspired me. Thank you to my family for putting up with my tyrannical moods and endless research agenda.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Frauds and Scams

Fraud is commonly described as the unlawful attainment of something of value realised through deceptive means (Hays and Prenzler, 2002, Lea et al., 2009, Stabek et al., 2009, and Wahlert, 1998). A scam is a tool adopted by fraudsters and used for agenda optimisation, thus a scammer is also a fraudster, and a scam therefore belongs to the family of fraud existing as a subset of fraud. In this section, a formal definition of fraud is introduced.

According to Hays and Prenzler (2002), a hierarchy of fraud is known, beginning with four categories of fraud which can be defined by; a) the intended target of the scam, b) the role of the perpetrator, and c) the method of scam perpetration (see Figure 1). To formally define these fraud categories let *A, B, C* and *D* represent each of the four fraud sets where *x* is the perpetrator, *y* represents the intended target of the scam and *z* is the method of scam perpetration:

A: *x* = a principle or senior official, *y* = an organisation, *z* = undefined

B: *x* = a client or employee, *y* = an organisation, *z* = undefined

C: x = undefined, y = a number of individuals, *z* = print or electronic media

D: *x* = an individual, *y* = an individual, *z* = face to face

Figure 1: Fraud Hierarchy

For fraud types *A* and *B*, only the role of the perpetrator and the target is known. For these two fraud clusters the target of the scam is faceless and the fraud is committed by someone with privileged access or knowledge of the target system or organisation. These two fraud types can be reconciled with the type of action that may be attributed to a person who may be a disgruntled employee or a group of individuals with a common agenda which may manifest as industrial espionage. In fraud cluster *C,* the role of the perpetrator is unknown, the target is a wide audience, and the method of perpetration is through the use of various types of media and technological communication systems, such as the Internet. For the final fraud cluster *D,* three variables are known: the perpetrator is an individual, the target is another individual, and the method of fraud perpetration is through face to face interaction. This type of fraud describes a situation where the perpetrator is known to the victim, the crime having evolved over the course of ongoing communication and developed relationships.

A scam is successful if it reaches its intended victim and, elicits the desired response from the receiver. Some receivers are aware of the signs of a scam and disregard the communication, while others respond to the psychological tactics used by the scammer (Lea et al, 2009). For every victim response there are hundreds and possibly thousands of non-responses and for this reason, one of the primary business processes (Lea et al., 2009) involved in a successful scam campaign is marketing, which involves the mass distribution of the scam to as many individuals as possible. For this reason, Internet distributed scams naturally lay within fraud cluster *C;* where the perpetrator is unknown, the target of the scam is a group of individuals, and the method of perpetration is through print or electronic media. While a hierarchy of fraud is demonstrated, commonalities among scams appear regularly and this subsequently makes it difficult to correctly and accurately identify scam perpetrations as the type of incident that they are (Hays and Prenzler, 2002, and Stabek et al., 2009).

A key point needs to be emphasised here about the nature of organised crime – these organisations often model themselves along the lines of traditional businesses, thus, it is usual for there to be a "Marketing Department" as well as other business units that fulfil core business functions (such as recruitment and financial management). In this sense, fraudsters organise their operations by designing and implementing business processes analogous to legitimate enterprises (Choo et al., 2009).

The focus of this research is on scams, rather than the broader body of fraud. Due to the hierarchical nature of the fraud ontology, at times the research crosses over from the cluster of fraud defined as fraud type *C* to the other fraud types and categories. The original focus for this research was technology based scams which encompassed Internet-assisted scams, however, through rigorous research and investigation it has been recognised that alternatively disseminated scams such as those described by fraud cluster *D*, technology may be used at some point throughout the life-cycle of the scam. Therefore, one of the main assumptions for this body of research is that all scams, regardless of their dissemination mechanics, incorporate the use of technology to facilitate perpetration.

The widely complex nature of scams, the intended targets, the methods of dissemination, the mechanics of scam success and the diverse nature of clientele falling victim to these fraudulent acts creates a challenge for industry and law enforcement alike. Mass communicability, cross border liaisons and speed of connectedness enables scammers to optimise their scams (Choo et al., 2008, and Wahlert, 1998). Due to the abundance of information available to individuals and instantaneous communication realised by the Wide World Web, scamming incidents have found their way into the global headlines (Drummond, 2010, United States Department of Justice and Federal Bureau of Investigation 2010). Since a scammer can operate many scams from one location, on multiple victims, in several countries, from numerous jurisdictions in unison, communication and cooperation between enforcement and investigative agencies across national and international borders is necessary for investigating and combating these crimes (Choo et al., 2008, Stabek et al., 2009, and Wahlert, 1998).

To facilitate police investigations and propel cross border liaisons into a transnational state of communicability and cooperation, standardisation of scam language consisting of scam labels, descriptions and scam definitions is necessary (Choo et al., 2008, Stabek et al., 2009, and Wahlert, 1998). Such a catalogue must achieve two outcomes: it must identify the actors and processes involved in "running the business" of scamming, as well as map standardised descriptors to those currently used in different jurisdictions. By adopting a common language of scams, inter-agency and cross-agency networking would be strengthened, leading towards cooperative cross-jurisdictional efforts in identifying, tracking, intercepting and prosecuting scammers.

Part of the problem faced by researchers of scam events is that there are numerous variations of the same scam all of which share interchangeable labels. This makes it very difficult for investigators, victims, and families of victims to confidently identify what they are, or have become involved in. Below are some examples of common scams received by email which illustrate this variation, and underline the need for a consistent approach to catalogue and describe the underlying processes involved in running a scam-based business.

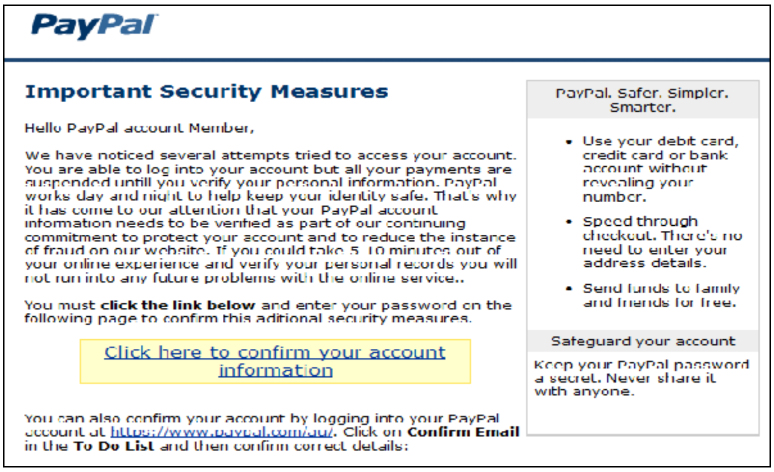
The first example seen in Figure 2 is a scare spam campaign which aims to panic the receiver into forwarding the communication onto everyone in their address book. This type of scam is commonly called a 'chain mail' scam and it requires the receiver to act impulsively by forwarding the scam on without fully considering the possibility that it is fraudulent. This scam uses the names and titles of reputable organisations to lend it some authenticity. The purpose of this type of scam is to find its way into as many email in-boxes as possible. It could be an information gathering campaign or it could contain malicious content such as a virus or spyware which would be downloaded onto the receiver's machine without their knowledge.

```
"HUGE VIRUS COMING ! PLEASE READ & FORWARD !

Hi All,

I checked with Norton Anti-Virus, and they are gearing up for this virus!
I checked Snopes, and it is for real. Get this E-mail message sent around to
all your contacts ASAP.

PLEASE FORWARD THIS WARNING AMONG YOUR FRIENDS, FAMILY AND CONTACTS!

You should be alert during the next few days. Do not open any message with
an attachment entitled 'POSTCARD FROM HALLMARK,'regardless of who sent it to
you. It is a virus which opens A POSTCARD IMAGE, which 'burns' the whole
hard disc C drive of your computer.

This virus will be received from someone who has your e-mail address on
his/her contact list. That is the reason why you need to send this e-mail to
all your contacts. It is better to receive this message 25 times than to
receive the virus and open it!

If you receive a mail called' POSTCARD,' even if it is sent to you by a
friend, do not open it! Shut down your computer immediately. This is the
worst virus announced by CNN.

It has20been classified by Microsoft as the most destructive virus ever.
This virus was discovered by McAfee yesterday, and there is no repair yet
for this kind of virus. This virus simply destroys the Zero Sector of the
Hard Disc, where the vital information is kept.
COPY THIS E-MAIL, AND SEND IT TO YOUR FRIENDS.
REMEMBER: IF YOU SEND IT TO THEM, YOU WILL BENEFIT ALL OF US
```

Figure 2: Scare Spam[1]

[1] All scam examples were delivered to the authors e-mail in-box and are copied verbatim from the original e-mail received

The PayPal email below in Figure 3 could be called a 'spam scam', however, by clicking on the provided link the receiver is redirected to a falsified or spoofed webpage where he/she are encouraged to divulge their personal information. By providing their personal details in this manner, the receiver would become involved in a scam called 'phishing'; this scam then becomes a spam and phishing scam. The scammers have taken measures to ensure that the scam looks authentic by using formal language and the appropriate logos of the target company. Similar to other phishing tactics, this scam uses fear and urgency to entice the receiver to act before they logically process the request.


PayPal

Important Security Measures

Hello PayPal account Member,

We have noticed several attempts tried to access your account. You are able to log into your account but all your payments are suspended untill you verify your personal information. PayPal works day and night to help keep your identity safe. That's why it has come to our attention that your PayPal account information needs to be verified as part of our continuing commitment to protect your account and to reduce the instance of fraud on our website. If you could take 5-10 minutes out of your online experience and verify your personal records you will not run into any future problems with the online service..

You must **click the link below** and enter your password on the following page to confirm this aditional security measures.

Click here to confirm your account information

You can also confirm your account by logging into your PayPal account at https://www.paypal.com/au/. Click on **Confirm Email** in the **To Do List** and then confirm correct details:

PayPal. Safer. Simpler. Smarter.

- Use your debit card, credit card or bank account without revealing your number.
- Speed through checkout. There's no need to enter your address details.
- Send funds to family and friends for free.

Safeguard your account

Keep your PayPal password a secret. Never share it with anyone.

Figure 3: Spam Scam

The following example in Figure 4 is a 419 Spam Scam, this example reads like it is an addition to a long line of communications and since it was unsolicited, is a spam scam. This email is initially a 'spam scam' since it was unsolicited. It also carries signs of a scam known as 'Nigerian Letter' or '419' with the seduction of a large sum of money and the use of official sounding people and places to add credibility to the communication. The receiver would be tricked into thinking that they had accidentally intercepted a communication and that they might be able to assist in the proposed transaction, for the identified fee. If the receiver were to contact the named associate and provide their bank account information, a case of identity theft would be probable leading to identity fraud and the subsequent draining of the victim's bank accounts.


I have a new email address!

You can now email me at: **kobellodebor001@att.net**

- Dear Good Friend,Hope you are doing good with your family today.It is my pleasure to let you know about my success in getting those funds transferred under the cooperation of a new partner from Greece.However,i didn't forget your past efforts to assist me in transferring those funds.Now contact my secretary Mr.Alan Garcia, his personal e-mail address (alangarcia21165@gmail.com)Ask him to release to you the total sum of One Million Four Hundred Thousand United States Of America Dollars (US$1,400,000.00) Certifie d Bank Draft which i raised for your COMPENSATION Due to your efforts during the course of the transaction.The amount could also be transferred to any of nominated bank account by bank wire transfer.So feel free and get in touch with Mr. Alan Garcia,he is REPUBLIC OF TRINIDAD AND TOBACO,because i have instructed him I'm very busy here with investment projects which i am having at hand with my new partner.Let me know immediately you receive your draft check or by in your bank account.Warmest Regards,Mr Kobello Debor.

Figure 4: Nigerian 419 Spam Scam

Below in Figure 5 is another phishing style scam, which could also be a syntactically driven spam campaign. It requires the recipient to click on the supplied link, at which time they would be redirected to a falsified or spoofed Webpage where they would be encouraged so supply their personal details, or, the link might contain malicious code which would download malware to the user's machine. A common theme emerges throughout all phishing–style tactics and that is authenticity. This example contains the Australian Coat of Arms and logo for the Australian Taxation Office (ATO). The timing of receipt for this scam was also pertinent to the success of the scam. This was distributed within 30 days from the end of the Australian financial year, when those who had completed their tax returns would be expecting to receive a response from the ATO.

Figure 5: Phishing Spam Scam

This last example seen in Figure 6 exemplifies the opening communication for a possible romance scam. It uses paraphernalia such as the attachment of a photograph as well as broken English to humanise the author. It contains a story of dreams and desires and briefly details the hopelessness that is felt by the woman pictured (supposedly). By responding to such a scam, the recipient would become involved in a romance scam which could transpire over many months ending with the payment of fees and bribes to the author's 'homeland officials'. For this type of scam campaign, a spam scam has been received and the victim may be urged to supply not only financial assistance, but their personal and private details. In which case, identity theft could occur.

Hello!

It is my first letter on English. Sorry, if I made some mistake in words. But I write you from my hand and don't use pre-written letters.
I am very glad, that you have become interested in me. And I shall try, that you were not disappointed with me and have learned as much as possible about me.
But I would like to learn you better too. I will ask you, write to me more about you in details.
My name Ekaterina. I live in Ukraine, in city Odessa. I am 28 years old.
If you think, that I am not serious don't make mistake, and know me much more. I gave promise, that I will never married on Russia boy.
All of them lie and don't hold his word. Some man drink alcohol very much. May be I will tell you more about my past relation later.
But i don't like think about it, it was no good.
My family are not large. We live with my mother. My mother have good work as bookkeeper. We can pay for all life expenses.
And I will not ask you help me with money. I know many stories about it. If you will write to me more, you will understand, that I am not such girl!
I want write to you long letter with much ideas from me, but I think, It will not good for the first letter. I am simple Ukraine girl, who want to live abroad.
I want have husband and right family. I will try for this very much. I have very serious intention.
My girlfriend find her husband on internet in last year. She move to Australia and they have happy family.
She lives in Sydney and they will have child soon. She write to me letter every week.

I was glad for it very much. We want to meet some time soon. I have great opportunity move to Australia at the end of this year.
Don't want write about me and my hobby in first letter.
We can talk on the phone, if you will want it. I don't have own phone, but I can use one from my friend or I can use call servise on post.
I will glad, if we can chat on MSN. I stay there at the evening and we can talk about all.
Please, send to me your phone number or MSN name, if you want contact with me from other way.
If You really interested in me, you can ask me about all.
I want ask you some question:
Do you have children? What are you doing at work? Did you have past relation, wife?
I hope, you can know some new things about me from this letter.
You can write to me on my e-mail: suhorukovakat@gmail.com
I will wait your letter and hope to receive news from you shortly.
Good luck to us.
Suhorukova Ekaterina.

Figure 6: Romance Spam Scam

Those scam examples presented here are only a small sample of the scams circulating the Web. Scams are not limited to the Internet however; they can occur in person like with 'door to door' scams. A victim could be recruited by a friend as seen with 'pyramid' and 'Ponzi' style scams, with many scams still perpetrated in more traditional forms such as over the telephone, and through the post. In most situations, the victim of a scam has responds to something that they have received and in other instances, the victim may have inadvertently sought out the scam which is a feature of 'ticketing' scams and some Internet auction scams. With the breadth and diversity of scams and scammer tactics increasing, it is imperative for research to focus on methods of identifying and combating these crimes (Airioldi and Malin, 2004, ABS, 2007, ACPR and AUSTRAC, 2006, ARC, 2009, Birzer and Craig-Mooreland, 2008, Choi, 2008, Denman, et al., 2004, Dolan, 2004, Goode et al., 2008, Hays and Prenzler, 2002, Jie et al., 2004, Lea et al., 2009, OFT, 2006, Stabek et al, 2009, and Wahlert, 1998).

Having reviewed a set of representative scams, some of the key issues are clear for identifying these for law enforcement purposes: (a) the "data" of the scam is natural language text, (b) the scam text describes the current state of some (business) process which require further action from the recipient; given the variation in text descriptions, trying to categorise these scams as belonging to a particular "group" could be quite challenging for individual investigators who each have their own biases, jurisdictions etc that will influence their decisions. What is required is a process that will allow the scam business processes to be identified from the text descriptions in each scam communication, and match these to some agreed template for a specific type of scam. Potentially, new e-mails and websites could then be classified in real-time and users alerted that a message may be a scam before they are tempted to click on a link that might take them to a phishing website, for example. In addition, law enforcement could use such a system to aggregate and identify common scams linking scammers and build a case against them.

A prerequisite to identifying scams is the identification of common business processes for scam types across jurisdictions. The first goal of the research presented in this thesis is to propose a technique for developing these templates in an attempt to objectively identify homogeneous groups of scams that are derived from text-based descriptions of scam types from a number of jurisdictions. The second goal is to then use the commonalities between the scam descriptions to identify hierarchical relationships between scam types, based objectively on their business descriptions, rather than *a priori* notions of what scams might be related to others. For example, the terms "identity theft" and "identity fraud" are often used interchangeably, but when you look at the commonalities between independent descriptions of their business processes, are they the same? Or, does "identity theft" have more in common with other types of "theft" rather than "fraud"? These are the kinds of issues that having a technique to objectively identity homogeneity and hierarchy within business process descriptions could help in resolving.

In this thesis, homogeneity and hierarchy are established by using a vector space model to represent "static" business elements derived from text descriptions of scams sourced from multiple authorities and then using hierarchical cluster analysis to group the scams and quantify their relatedness. In addition, approaches for model validation are also introduced. By deriving data from publicly available scam descriptions, hierarchical clustering will ensure homogeneous partitioning of scams into similar clusters that can be inferred from scam static features. This forms analogous scam clusters which can then be used for building the sorts of templates that could be matched from potential scam materials, such as e-mails and websites. Secondary to the clustering of scam cases by their descriptions is the standardisation of scam descriptions and identification of significant static features which can then be used to confidently identify the type of scam a scam is.

Furthermore, reduction of scam events into homogeneous scam clusters will assist investigative and law enforcement agencies by reducing time, money and resources spent on scam case investigations. It is also hoped that the results from this research will lead the way towards a common scam lexicon and enhanced coordination and cooperation between transnational taskforces.

More generally, the approach could potentially be generalised to other types of business process modelling, where there are no formal descriptions of processes marked up in a language, such as BPXML. A series of candidate classes and their relationships, representing static data elements, could be inferred, although it is beyond the scope of this thesis to pursue applications of the approach outside cybercrime. The set of representative scams reviewed here provide evidence of some of the commonalities within scam-types and the examples discussed within this chapter verify these commonalities across scam categories. The reduction of scam events into homogeneous scam clusters will assist investigative and law enforcement agencies as well as help to standardise the comprehension of scam-types amongst reporting institutions. This in turn will assist in the useful comparison of scam incidence across jurisdictions which would provide more accurate and precise results. The following section expands on the methods and approaches used by international and national agencies reporting on the incidence of scams and frauds.

1.2 Scams and Statistics

This research was originally influenced by the investigation of publicly available annual reports on scam incidents. Four main contributors originating from four different countries were investigated; the Internet Crime Complaint Center (United States of America), the Australian Bureau of Statistics (Australia), the Environics Research Group (Canada) and the Office of Fair Trading (United Kingdom). From the initial investigation it was recognised that the data collection, analysis and identification of scam events across reporting agencies was inconsistent. Presented below, is an overview of these challenges and an outline of the methodologies used by each reporting agency.

1.2.1 Internet Crime Complaint Center

In 2001 the Internet Fraud Complaint Center (IFCC), in collaboration with the National White Collar Crime Center (NWC3) and the Federal Bureau of Investigation (FBI) joined forces to produce the first ever annual Internet Fraud Report (IFCC, 2001). The IFCC operates to this day as the Internet Crime Complaint Center (IC3) and receives complaints in the form of victim self-report data pertaining to Internet and computer based crimes. In 2001 the IFCC received 49,711 complaints and referred 34% of these on for further inquiry (IFCC, 2001). Within the following 12 month period, the number of victim complaints grew to 75,063 with a 64% referral rate (IFCC, 2002) and by 2008 a total of 275,284 complaints were received (IC3, 2008).

Since the launch of the iC3 in 2001, email has remained the most optimal method of distributing scamming material to potential victims (see Table 1, below). The percentage of received email-based scamming communications has increased by 6.4% between 2001 and 2008 and this method of disbursement accounts for almost 50% (48.5) of all reported scamming communications for 2008. Web page-based scamming distributions were reported in 28.9% of all victim reports which represents an increase of 12.5% since 2001. In 2005, web page-based proliferation was reported in 16.5% of cases which increased to 36% the following year. This almost doubling of web page-based incidents may be explained by the change in detail of what is actually reported upon by iC3. It appears that during the years 2001 to 2005, the iC3 reported the percentages for ‘method’ for only those cases referred on for further investigation while the following years, 2006 to 2008, the method for the total number of complaints received appears to be reported. The third reported most utilised method of scamming distribution was by phone which accounted for 15.6% of all received complaints in 2008. Up until 2005, the category of ‘printed material’ was ill-defined which is apparent by the development of new, more precise categories in the following years. These developments were the identification of ‘newsgroup’, ‘bulletin board’, ‘wire’, and ‘instant messenger’ which would take the place of the category known as ‘printed material’. Interestingly, scammer reliance on printed material for the distribution of their schemes seems to have increased for the years 2006 to 2008 since the use of printed materials reached an all time low in 2005 with less than 1% of complaints, while in 2006, 2007, and 2008 this increased to 22.6%, 21.2% and 18.8% respectively.

The percentage referral rate received by the iC3 has an impact on the reported statistics for total dollar losses of each year which are presented in Table 2. Based on those complaints which were referred on for further investigation and the monetary loss associated with these referrals, the total monetary loss in the years following 2001 increased by $US107.8 million with the average loss increasing by $US926.60. During 2003, there was a $US71.6 million decrease in total funds lost and since then, there has been a steady increase in dollars lost. In 2008, a $US264.4 million loss was recorded with an average loss of $US3,637.70 and a median loss of $US931.00.

Table 1: Complaint Percentage by scamming method

Percentage of complaints by method of scam introduction for each year								
Method	2001	2002	2003	2004	2005	2006	2007	2008
Email	68.4	66	64.8	63.5	73.2	73.9	73.6	74
Web page	13.4	18.7	19	23.5	16.5	36	32.7	28.9
Phone	9.6	7.6	8	7	4.5	17.7	18	15.6
Post	4.2	3.9	4.1	3.2	2	10.3	10.1	8.3
Printed material	1.9	1.7	1.4	1.2	0.9	0	0	0
In person	1	1	0.9	0.6	0.8	1.5	1.7	1.7
Chat room	0.8	0.7	1	0.7	1.8	2.4	2.3	2.2
Fax	0.8	0.4	0.5	0.2	0.3	4	3.5	3.1
Newsgroup	0	0	0	0	0	0.6	0.5	0.5
Bulletin board	0	0	0	0	0	3.7	3.9	3.8
Wire	0	0	0	0	0	6.3	5.3	4.2
Instant messenger	0	0	0	0	0	12	11.5	10.3

Table 2: IC3 Recorded Dollar Losses to Internet Crime

Year	Total $US million	Average $US	Median $US
2001	$17.80	$1,061.10	$435.00
2002	$125.60	$1,983.70	$329.00
2003	$54	$1,119.13	$299.00
2004	$68.14	$655.40	$219.56
2005	$183.12	$1,886.40	$424.00
2006	$198.44	$2,300.00	$724.00
2007	$239.09	$2,656.32	$680.00
2008	$264.60	$3,637.70	$931.00

Table 3: IC3 Top Ten Recorded Internet Crime Scams

Top 10	2001	2002	2003	2004	2005	2006	2007	2008
1	Auction fraud	Auction fraud	Auction fraud	Auction fraud	Auction fraud	Auction fraud	Auction fraud	Non delivery
2	Non delivery	Non delivery	Non delivery	Non delivery	Non delivery	Non delivery	Non delivery	Auction fraud
3	Nigerian letter fraud	Credit/debit card fraud	Nigerian letter fraud	Credit/debit card fraud	Credit/debit card fraud	Check fraud	Confidence fraud	Credit/debit card fraud
4	Credit/debit card fraud	Investment fraud	Credit/debit card fraud	Check fraud	Check fraud	Credit/debit card fraud	Credit/debit card fraud	Confidence fraud
5	Confidence fraud	Business fraud	Confidence fraud	Investment fraud	Investment fraud	Computer fraud	Check fraud	Computer fraud
6	Investment fraud	Confidence fraud	Investment fraud	Confidence fraud	Computer fraud	Confidence fraud	Computer fraud	Check fraud
7	Business fraud	Identity theft	Business fraud	Identity theft	Confidence fraud	Financial institutions fraud	Identity theft	Nigerian letter fraud
8	Identity theft	Check fraud	Identity theft	Computer fraud	Identity theft	Identity theft	Financial institutions fraud	Identity theft
9	Check fraud	Nigerian letter fraud	Check fraud	Nigerian letter scam	Financial institutions fraud	Investment fraud	Threat	Financial institutions fraud
10	Communications fraud	Communications fraud	Intellectual property fraud	Financial institutions fraud	Child pornography	Child pornography	Nigerian letter fraud	Threat

Table 3 above lists the top ten scams for each year in descending order. With the exception of 2008, auction fraud was consistently the most reported scam followed by the non-delivery of merchandise scam. The consistency of appearing in the top ten lists for those reported scams suggests that the identified scams are stable over time. With this historical perspective of scam events within the USA, it can be concluded that non-delivery, auction fraud, credit/debit card fraud, confidence fraud, computer fraud, check fraud, Nigerian 419 fraud, identity theft, financial institutions fraud, and threats, are likely to be problematic in ensuing years.

1.2.2 Australian Bureau of Statistics

From niche focussed reporting institutions such as the iC3 to broadly defined information collection and analysis agencies such as the Australian Bureau of Statistics (ABS), the impact of scams are felt at all levels of society. The ABS released its first ever Personal Fraud Report (PFR) during 2008. The report details the results of a telephone survey which was conducted as an addition to the Multi-Purpose Household Survey (MPHS). The ABSPFR (2008) achieved a sample of 14,320 participants reporting on their experiences and losses associated with the loosely defined concept of personal fraud. It was reported that financial losses attributable to personal fraud were in excess of $AUS980 million. The concept of 'personal fraud' remained undefined and it is assumed that two exclusive categories of crime lead a victim to the exposure of personal fraud; identity fraud and scams. The ABSPFR (2008) collected information pertaining to participant experience in receiving and responding to possible fraudulent and scam-based invitations. The events identified as identity fraud crimes were credit or bank card fraud and identity theft while those events identified as scams were lotteries, pyramid scams, phishing, financial advice scams, chain letters, advance fee fraud and all other scams (ABS, 2008).

The ABSPFR (ABS, 2008) results suggest that for the year leading up to the survey, 5.8 million Australians were exposed to a scam with 1.9 million of those earning less that $AUS499 per week and 208,000 people earning on average $AUS2500.00 per week. For those Australians exposed to a scam, 329,000 were victimised by a scammer. The ABSPFR (2008) advise that a person could be exposed to a scam without becoming a victim; however, if a person were exposed to an identity fraud incident, they immediately became a victim. There were 882,800 victims of personal fraud and 124,000 of those were victims of identity theft (ABSPFR 2008).

Those scams which achieved the greatest amount of public awareness were lotteries (2,437,400), phishing (2,374,700), and chain letters (2,054,000) while those scams achieving the greatest number of victims were lotteries (0.5%), pyramid schemes (0.4%), and phishing (0.4%). Respondents to the survey reported receiving lottery scam invitations most of the time over the Internet via email or by post (342,000 and 330,000), while pyramid scheme invitations were most often received in person (419,000). Phishing (and related) scam requests were disseminated most regularly via the Internet and email (304,000) followed by telephone (157,000). Financial advice scams were reported to have only occurred by two channels, these were face to face communication (193,000) and email (92,000) while chain letter scams were only recorded as having been received by email (129,000) and post (138,000). The method of scam introduction was not recorded for those recipients of advance fee scams while the methods of distributions for all other non-defined scams were phone (194,000), email (317,000), post (90,000) and 'other' which includes face to face interactions (90,000). The method of scam distribution is a clear indicator of scam type and thus is a static feature of scam construction which will be discussed in further detail in subsequent chapters.

1.2.3 Environics Research Group

The 2007 Consumer Mass Marketing Fraud Survey compiled by the Environics Research Group (ERG) based out of Canada defined Mass Marketing Fraud (MMF) as mass communication network-assisted fraud (ERG, 2008). There were 12 identified scams categorised as MMF which were; prize, lottery or sweepstakes fraud, West African or 419 fraud, employment/work from home fraud, cheque cashing/money transfer job fraud, overpayment for sale of merchandise fraud, advance fee loan fraud, upfront fee for credit card fraud, bill for unsuitable merchandise fraud, bogus health product or cure fraud, advance fee vacation fraud, high pressure sales pitch vacation fraud, and investment fraud (ERG, 2008).

Similar to the ABSPFS (2008), the MMF survey was administered by telephone interview and only included Canadian consumers, excluding Canadian businesses. It was reported that 15 million Canadians became victims of scams losing $CAD450 million during 2007 and it was identified that 9 in 10 victims of scams do not report their experience to the authorities. It is reported that while recipients of scamming communications change their consumer behaviour in over half of the interviewed cases, this does not reduce the likelihood of being targeted; survey respondents received on average 16 scam proposals per year which reportedly increased by 31% for those who had responded to a communication in the past (ERG, 2008). The ERG indicates that scams traverse all demographic and socio-economic channels, implying that any one person can become a victim of mass marketed scams.

1.2.4 Office of Fair Trading

The 2006 the United Kingdom Office of Fair Trading (UKOFT) Research on the Impact of Mass Marketed Scams (MMS) (OFT, 2006) estimated that the total UK dollar loss to MMS was £3.5 billion per annum, averaging £70 each year per adult residing in the UK. The report identified 15 categories of scam type and suggested that the average dollar loss per scam type was £850. Those scams identified as MMS were; prize draw/sweepstakes scams, foreign lottery scams, work at home and business opportunity scams, premium rate telephone prize scams, miracle health and slimming cure scams, African advance fee frauds/foreign money making scams, clairvoyant/psychic mailing scams, property investor scams, pyramid selling and chain letter scams, bogus holiday club scams, Internet dialer scams, career opportunity scams, high risk investment scams, internet matrix scams and loan scams (OFT, 2006).

Consistent with the findings of the ERG, the UKOFT recognises that the victims of scammers are not identifiable by their age, gender, socio-economic stats or educational background. While it is recognised that gender is not significant to victim status, gender does play a role in the type of scam that victims might fall for. An example of this is the admission that women were most enticed into victim status by miracle health scams (71%) while men were most likely to fall victim to high risk investment scams (72%). Similarly, while socio-economic status is independent of scam victimology, it is a factor in the type of scam that victims of lower, middle or high social class fall for. Low income earners and the working class were affected mostly by loan, foreign lottery, career opportunity and clairvoyant scams while the middle to upper class were more likely to become victims of African advance fee, property investment and high risk investment scams. A clear separation between the class of the recipient and the type of scam that they fall victim to is apparent. In the case of low income earners and the working class, they are most likely to fall victim to those scams that offer opportunity such as employment and prizes while the middle to high income earners are most likely to fall victim to investment style scams where they need to outlay an higher amount of money up front. It is also worth noting that while equality of rights movements have advanced the social standing of women in the workforce, the majority of women sit in the working class to middle band of social status, this may be why for this representative sample those scams exploiting female and male tendencies are also separated by social standing.

The methods of data collection used by the UKOFT (2006) were focus discussions, in depth interviews, omnibus surveys, and telephone interviews. The results suggest that approximately 48% of the UK populous had been targeted by a scam in the past and an estimated 3.2 million people would become tricked by scammers on average per year. The separation of those people who were targets of scams compared to those people who were became victims is similar to that of the ABSPFR (2008) pertaining to scam exposure and scam victimisation. A challenge faced by researchers was the admission or realisation by participants that they may have become a victim of a scam while it is reported that in less than 5% of cases, a scam will be reported to the authorities.

In 2009, the UKOFT released an inquiry into the Psychology of Scams (Lea et al., 2009). The report identified similarities between legitimate product marketing techniques and illegitimate product marketing approaches suggesting that scammers adopt common processes models for illegitimate product exploitation. Similarly, Choo et al (2009) propose the criminological use of standardised business processes in the development and implementation of scamming campaigns. A mixed methodology involving four approaches; in depth interviews, text mining, questionnaires, and behavioural experimentation were used.

The report documented and analysed the psychological principles involved in 10 scams: advance fee 419 scams, international sweepstake claims, fake clairvoyant scams, prize draw pitch scams, get rich quick scams, bogus investment scams, bogus lottery scams, miracle health cure scams, premium rate prize draw scams, and bogus racing tipster scams (Lea et al., 2009). Identifying the primary psychological tactics employed by scammers as the triggering of visceral processes, norm activation, perception of authority, reduction of motivation for information processing, and liking and similarity. Lea et al. (2009) suggest that visceral processes are activated by a reference to a large reward – such as those demonstrated by lottery and prize scams, norm activation uses the exploitation of human desires to achieve a desired response and outcome - such as the desire to work from home and make a lot of money with little outlay as with employment or investment scams, fear or urgency result in the diminished activation of information processing – as seen with phishing and death threat scams, while liking and similarity refer to the relatable ability of the scammer to empathise with the victim – common to Nigerian letter and romance scams.

The results and approaches used by different reporting agencies from various institutions and numerous countries have been presented above. These reports suggest that public, social and industry awareness of scams is growing and they also dispel some myths as to the likely targets and majority of victims of scams. Each reporting institution utilised their own definitions and criteria for examining the incidence of scams in their country and jurisdictions with varying foci for their investigations. The competing terminologies and mixed understandings of the language used to describe scamming events amount to discrepancies in scam report data both within jurisdictions and across national borders. The borderless nature of these crimes is an indication of the type of approach that is needed to combat the issue. While the authorities and reporting agencies continue to work autonomously and independently little progress can be made towards combating these crimes. Below is a more in depth examination of the methodologies adopted by each of the reporting institutions detailed above and a case is developed for the necessity of reform.

1.2.5 Approach Comparison and Description Discrepancies

A comparison of the methods used by each of the four reporting institutions detailed above; the ABS, ERG, iC3 and OFT is presented below in Table 4 along with the number of scams identified by each report. It is shown that both the ABS and ERG used telephone interviews for the collection of data, the iC3 collated data from victim self reports while the OFT used a variety of methods to collect data; focus group, in depth interview, survey and telephone interview. The number of identified scams associated with each group is different for each reporting institution and is a reflection of the priority focus of each institution. The ABS was only concerned with 'Personal Fraud' and reported upon 9 different scam types. The ERG was concerned with mass marketed scams and identified 12 different scams within this category. The iC3 receives reports on all types of computer crime while reporting upon those scams that are referred onto other law enforcement authorities for further investigation and as such only the tope ten scams for the annum are discussed. The OFT released two separate reports; one on mass marketed scams and one on the psychology of scams, these two reports combined identified 25 scams.

Table 4: Method Comparison by Reporting Institution

Insititution	Methods	No. Scams
ABS	Telephone survey	9
ERG	Telephone survey	12
Ic3	Victim self report	10
OFT	Focus group, in depth interview, survey, telephone interview	25

The results of the iC3, ABS, OFT and ERG pertain only to the population from which they were sampled. For this reason, along with the different scam-type foci apparent between reporting groups it is difficult to make cross-border comparisons or acquire a true representation of the impact that scams have on business, law enforcement, the individual, family, and the community in each country that they belong. These barriers make it challenging to interpret the results in a real world or useable context. Further to these interpretation inconsistencies, the complexity and fluidity of scams is not represented in any of these reports which are often heavily relied upon as a point of reference by state and federal agencies.

Scammers benefit from a borderless advantage while investigative and enforcement agencies are stifled by the cross-jurisdictional nature of scams. To monitor, intercept and prosecute scammers it is necessary for cross-border liaisons to be strengthened and for the enhancement of transnational efforts which cannot be achieved without a common language of scams (Stabek et al., 2009). Some of these problems manifest in similarity amongst scam descriptions, as well as over simplified scam descriptions.

"Confidence Fraud results in the financial loss of an individual who has been swayed to act to their detriment through a breach of trust in a relationship. "

"Credit / Debit Card Fraud occurs when there has been unauthorised use of a credit card."

"Computer Fraud is the violation of law involving a computer. "

"Identity Theft is the unauthorised use of another's particulates"

(IC3, 2008)

Figure 7: IC3 Scam Definitions

The above examples given in Figure 7 demonstrate some of the similarities in scam definitions for the IC3 alone. The category of Confidence Fraud could imply all forms of scam since all scams contain the crucial element; *breach in a relationship* or transaction. Credit/Debit Card fraud could easily be confused with Identity Theft since Identity Theft is only defined as the unauthorised use of another persons details and the unauthorised use of another persons credit or debit card is *the unauthorised use of another's particulates*. And finally, since Computer Fraud is defined as the *violation of law involving a computer*, and all Internet and computer based scams involve the use of a computer in their perpetration, all scams and frauds perpetrated through the use of such technologies can be described by this one definition.

"Advance Fee Loan Fraud occurs when a loan is offered in return for the payment of an advance fee."

"Upfront Fee for Credit Card Fraud occurs when a credit card is offered in return for the payment of an advance fee."

"Advance Fee Vacation Fraud occurs when a discounted vacation is offered in return for the payment of an advance fee."

"Prize Lottery or Sweepstakes Fraud occurs when a prize is offered in return for the payment of an advance fee."

(ERG, 2008)

Figure 8: ERG Scam Definitions

The above examples in Figure 8 demonstrate the issue of over-classification. Advance Fee Loan Fraud, Upfront Fee for Credit Card Fraud, Advance Fee Vacation Fraud, and Prize Lottery or Sweepstakes Fraud contain a shared characteristic which underpins the definitions of the scams, this is something in return *for the payment of an advance fee*. Based upon the definitions given, these scams could easily be combined into one category of MMF rather than separated into four separate groups.

Table 5 below lists all of the scam titles identified by each scam reporting institution discussed in this chapter. Of the 56 scams reported upon, only one was recognised by an identical title by another institution. 'Identity theft' was labelled 'identity theft' by both the ABS and iC3 while no other scam was recognised by a same title between any of the four institutions or 5 individual reports. Those scam types that are alike such as 'Lotteries' (ABS), 'Prize, lottery or sweepstakes' (ERG), 'Prize draw sweepstakes' (OFT2), and 'Prize draw pitch' (OFT2) but receive competing labelling dependent upon institution occur frequently throughout the sample. As demonstrated in the given example, there are even discrepancies evident between reports issued by the same institution (OFT1, OFT2). These inconsistencies both within reporting institution – such as that demonstrated by the OFT, and between reporting institutions impedes current communication channels across national and international borders. Greater consistency in scamming language would assist towards the development of a comprehensive scamming lexicon that could be used to assist not only reporting institutions but local and transnational efforts between law enforcement agencies, financial institutions, businesses, and investigative and security professionals. Greater consistency means greater understanding, greater understanding means greater awareness and a greater awareness means a greater ability to address the issue. These issues are addressed further by the research goals which are discussed in the following section.

Table 5: Comparison of Scam Titles and their Reporting Institution

Scam Title / Institution	ABS	ERG	iC3	OFT 2	OFT 1
Advance fee 419				✓	
International sweepstakes				✓	
Fake clairvoyant				✓	
Prize draw pitch				✓	
Get rich quick				✓	
Bogus investment				✓	
Bogus lottery				✓	
Miracle health cure				✓	
Premium rate prize draw				✓	
Bogus racing tipster				✓	
Prize draw / sweepstakes					✓
Foreign lottery					✓
Work at home & business opportunity					✓
Premium rate telephone prize scams					✓
Miracle health and slimming					✓
African advance fee frauds/foreign money making					✓
Clairvoyant/psychic mailing					✓
Property investor					✓
Pyramid selling and chain letter					✓
Bogus holiday club					✓
Internet dialer					✓
Career opportunity					✓
High risk investment					✓
Internet matrix					✓
Scams and loan scams					✓
Prize, lottery or sweepstakes		✓			
West African or 419		✓			
Employment/work from home		✓			
Cheque cashing/money transfer job		✓			
Overpayment for sale of merchandise		✓			
Advance fee loan		✓			
Upfront fee for credit card		✓			
Bill for unsuitable merchandise		✓			
Bogus health product or cure		✓			
Advance fee vacation		✓			
High pressure sales pitch vacation		✓			
Investment		✓			
Credit or bank card	✓				
Identity theft	✓		✓		
Lotteries	✓				
Pyramid	✓				
Phishing	✓				
Financial advice	✓				
Chain letters	✓				
Advance fee fraud	✓				
All other scams	✓				
Non delivery			✓		
Auction fraud			✓		
Credit / debit card			✓		
Confidence fraud			✓		
Computer fraud			✓		
Nigerian letter			✓		
Financial institutions			✓		
Threat			✓		

1.3 Research Goals

There are three goals for this research which all combine to achieve the result of the identification of hierarchy and homogeneous subsets of scam perpetrations.

Research Objective 1: To cluster scam descriptions by partitioning them into scam genres revealing homogeneous groups of scam cases.

Methodology Outline 1: This will be achieved by analysing the divisively derived static features of scam descriptions following the manual content analysis of publicly available scam descriptions. By means of agglomerative hierarchical clustering, scam cases will be analysed according to their divisively derived static feature composition and partitioned into relatively homogeneous clusters of like scam genres.

Research Objective 2: To measure the effectiveness of using static features in identifying scam genres, where a scam genre is composed of a scam cases with similar compositions of static features.

Methodology Outline 2: This will be achieved using a discriminant function analysis which will test the significance of each static feature on the placement of scam cases into scam genres, quantising the effectiveness of using the divisively derived static features for identifying scam genre membership.

Research Objective 3: To verify the selected hierarchical clustering model which produces the relatively homogeneous subset of scam genres and is at least 95% reliable. This will ensure that scam genre placement will be accurate at least 95% of the time.

Methodology Outline 3: This will be achieved by applying discriminant function analysis to verify the results found from the hierarchical cluster analysis. The discriminant function analysis will compare the scam-case cluster assignments from the hierarchical model with the predicted scam placement from the discriminant function model. The hierarchical clustering model with an error no greater than 5% and that clusters scam cases into relatively homogeneous genres will be identified as the best fitting approach for clustering scam cases from the their divisively derived static feature composition.

The ability to confidently identify scam genres by scam cases explicitly aligns with Australia's National Research Priorities as detailed in the Designated National Research Priority, Research Priority 4: Safeguarding Australia (ARC, 2009).

1.4 Statement of the Problem

With the global connectedness and mass communicability offered by the Internet and World Wide Web, a scammer can operate one scam in numerous countries, across several jurisdictions on countless victims, in unison, all whilst managing multiple projects in parallel. The complicated network of scam transactions, cross-jurisdictional technicalities and speed of online banking and wire transfers aids scammer success, which is realised by authorities with difficulty found in identifying, tracking and prosecuting scammers. Scam success stories are compounding (IC3, 2001-2008) and organised cyber-criminal groups and individuals disseminating their scams from single or multiple locations and defraud victims as far reaching as the connectedness of technology offers (Choo et al., 2008). The need for interagency communication and cooperation leading towards cross-jurisdictional, transnational investigative operations is necessary (Choo et al., 2008, Stabek et al., 2009, and Wahlert, 1998) to combat the pandemic of scam proliferation. The successfulness of scam campaigns, evidenced by the increasing number of victims and public awareness, due largely to the speed of connectedness and psychology-driven selling tactics employed by scammers suggests an adaptation towards a scam management style synonymous with successful business models (Choo et al., 2008, Hays and Prenzler, 2002 and Stabek et al., 2009).

Without a shared and common lexicon of scam labels and descriptions, communications between agencies and cross-border cooperation is stifled; the biggest hurdle to overcome in scam investigation is inconsistent scam classifications (Stabek at al, 2009). Without consistency, a uniform language of scams cannot be forged and without a common language, ambiguities leading to interpretational impedances are aiding scammers by enabling their scams in cross-jurisdictional and multinational platforms.

While there is little research to date to either confirm or deny that scams are developed following particular business principles Choo et al. (2008), and Stabek et al. (2009) observe that scam operations are fluid and adaptive to change confirming that organised cyber-criminal groups, individuals and scammers successfully operate their scams by following proven business models. Without a complete comprehension of the processes involved in scam transactions, investigators cannot quantifiably or empirically assess scam situations. Before assessment of scam incidents can progress to a heightened level of analysis, there needs to be compatible understanding of the scam event being investigated which involves uniformity in scam descriptions across investigative and reporting agencies, leading to a shared understanding of scam events across jurisdictions. To the knowledge of this researcher, no method of scam description standardisation exists.

The problem with a naturally aligning language of scams is that there is too much diversity in recognised scam types across countries, borders and jurisdictions. This research aims to condense identifiable and publicly recognised scam events into homogeneous clusters of scam genres which will assist in the identification of scam categories. This will contribute to the development of a uniform lexicon of scam terminology as well as become foundational to further research on the impact of scams at a local and global level.

1.5 Research Problems

This research addresses three research problems. The first focuses on the richness of the source data and questions the usefulness of using scam static features in determining scam group membership. The second research problem addresses the issue of the method of classification. Due to the empirically devised fraud hierarchy, it is logically deduced that a hierarchy of scams also exists; therefore, this research component focuses on the efficacy of using hierarchical cluster analysis for the grouping of scam static features. The third and final research problem focuses on the validation of the hierarchical model produced from research problem two which is validated through the use of discriminant function analysis.

1.5.1 Homogeneous Grouping Problem

The homogeneous grouping problem explores the suitability of hierarchical cluster analysis in partitioning scam cases into scam genres. Four hierarchical linkage methods; furthest neighbour, between groups, within groups, and nearest neighbour, and two binary distance measures; Jaccard coefficient and Simple Matching coefficient have been identified for comparison in this research with a supplementary aim of identifying which combination of method of linkage and measure of distance best partitions scam descriptions into homogeneous groups and creates the 'best solution'. The concept of 'best solution' is been defined as a result which partitions scam cases into the fewest number of groups with the least number of scam cases allocated to each group, being therefore, relatively homogeneous.

1. Which binary linkage method (furthest neighbour, between groups, within groups, and nearest neighbour linkage) and binary distance measure (Jaccard or Simple Matching) best partitions scam descriptions into relatively homogeneous groups?

    a. Which cluster result contains the fewest number of groups with the least number of scam descriptions allocated to each group?

1.5. 2 Static Feature Selection Problem

The static feature selection problem explores the usefulness of scam static features in determining scam group membership and involves the determination of required static features bearing significantly on the placement of scam cases into scam genres.

2. Can static features be used to determine scam group membership of scam descriptions?

    a. How many static features are required to determine scam group membership?

    b. What static features are useful in determining scam group membership?

1.5.3 Minimum Cluster and Least Membership Problem

The minimum cluster and least membership problem considers the usefulness of discriminant function analysis in predicting and comparing scam group membership of the hierarchical clustering models. This problem requires the resultant conclusion to contain the fewest number of scam clusters and least number of scam cases in each cluster. A tolerance limit has been set for accepted amount of unexplained variation which is 5%. This means that for the selected clustering model, the best solution containing the least number of clusters and the fewest number of scam descriptions in each cluster should be able to accurately predict scam group membership at least 95% of the time.

3. Can a discriminant function analysis be used to predict scam group memberships for the hierarchical cluster solutions with the fewest number of clusters and least number of scam cases in each cluster to determine which solution accurately predicts scam group memberships at least 95% of the time?

1.6 Project Contributions and Chapter Summary

This research will contribute to scam and fraud literature as well as extend on current scam and fraud research methodologies. The methodology provided here offers a template for future cybercrime classification work and will be useful to the development of a National Cybercrime Reporting Service. Further to this, reduction of scam events into homogeneous scam clusters will assist investigative and enforcement agencies by reducing time, money and resources spent on scam case investigations. It is also hoped that the results from this research will lead the way towards a common scam lexicon and in turn enhanced coordination and cooperation in transnational taskforces.

In this introductory chapter evidence has been presented for the need to focus research on the incidents of scams and a case has been built urging for reform in the field of scam research. It has been demonstrated that there currently exists contradiction and inconsistency amongst scamming terminology both at institutional levels and across national borders. A consistent lexicon of scamming terminology has been identified as necessary in propelling research forward and enhancing communication, cooperation and understanding between law enforcement authorities, financial institutions and businesses which in turn would assist in the strengthening of transnational investigative efforts. The research agenda of this thesis has been outlined presenting the goals and research questions for this research. There are three primary research questions and areas of focus; 1) the usefulness of divisively derived static features in hierarchically clustering scam cases, 2) the identification of the best hierarchical model from a total of 8 different models which clusters scam cases relatively homogenously, and 3) the verification of hierarchical model which best clusters scam cases into relatively homogenous clusters, accurate at least 95% of the time. The rest of this thesis is organised into four sections; Literature Review, Methodology, Results and Discussion.

Chapter 2: Literature Review

2.1 Defining Scams

Lack of uniformity causing inconsistencies in scam classification has been identified as an ongoing concern since the 1990's (Wahlert, 1998) and is regularly identified as a primary cause of discrepancy in victim self report data (ACPR and AUSTRAC, 2006, Choo et al., 2008, Stabek et al., 2009). A recent article (Stabek et al., 2009) presents strong evidence citing the breadth of variation in scam classification among international and national scam reporting institutions. It is suggested that such wide inconsistencies cause reduced rates of scam identification and low rates of interception by law enforcement agencies, which results in low prosecution rates of scammers. Confusion, uncertainty, and false negatives are the consequences scam description discrepancies in scam classifications. These outcomes lead to more complex concerns, such as the under reporting of scam incidence, reduced rates of successful follow up by investigative and enforcement agencies and difficulty in making correct referrals (Denman et al, 2004, Dolan, 2004, Hays and Prenzler, 2002, Lea et al., 2009, and Wahlert, 1998). There are blurred boundaries within scam classifications which are inherent throughout anti-fraud legislation (Cybercrime Act, 2001) which the criminals seem to exploit to their advantage (Choo et al., 2008) and confusion and uncertainty also exists for transnational investigative bodies dealing with the complexities of working beyond home borders and interacting with multiple jurisdictions (Wahlert, 1998, and Stabek et al., 2009).

In this chapter scams are defined in terms of technology enhanced and technology enabled crimes. A modified model of technology based crimes is defined and it is suggested that this is necessary for the purposes of inclusionary research pedagogy in the field of scams investigation. An overview of scam-based research is delivered which is separated into subsections headed by the type of scam each research team focused on.

2.2 TEAC versus TEHC

According to the Australian Institute of Criminology (AIC) (Choo et al., 2008), there exist two types of technology based crimes, syntactic; crimes in which technology is the target of the scam, and, semantic; crimes in which the user is the target of the scam. Within the category of semantics, two groups of crime emerge; these are technology-enabled crimes (TEAC) and technology-enhanced crimes (TEHC), this distinction is represented figuratively below in Figure 9.

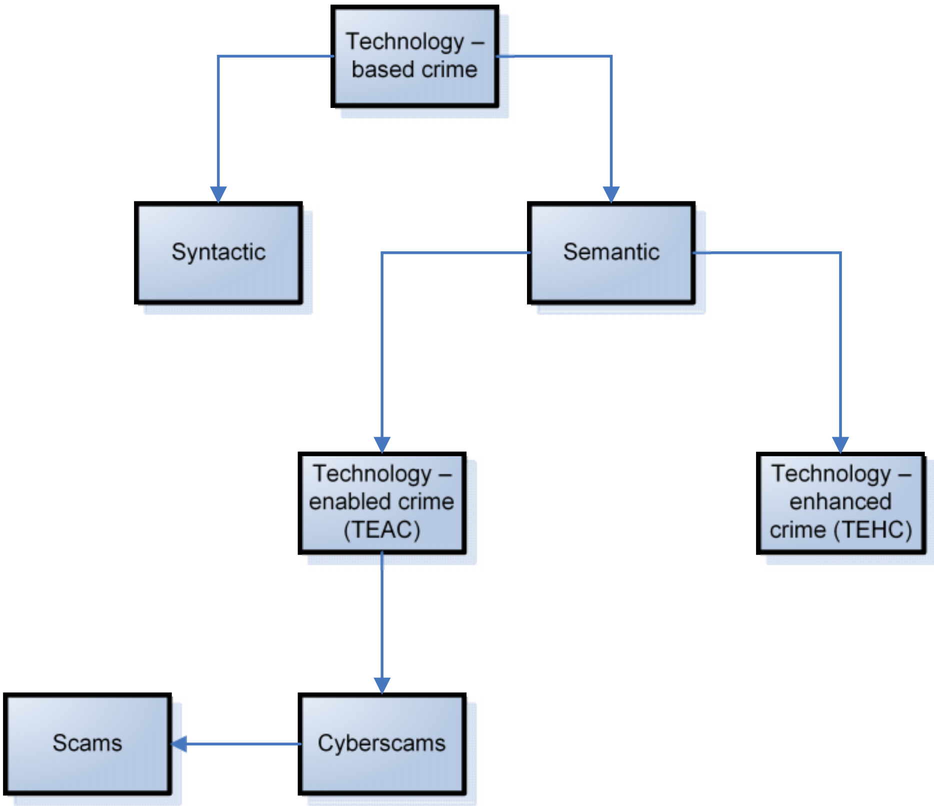


Figure 9: AIC Technology Based Crime Flow Chart

According to the AIC (Choo et al., 2008), a technology-enabled crime (TEAC) is a crime in which technology is necessary for its implementation, while a technology-enhanced crime (TEHC) is a crime where technology is used as a facilitator of the crime. A distinction can be drawn between technology required for the commission of crimes; TEAC, and technology useful in the implementation of crimes; TEHC. Use of technology is necessary for TEAC, while the use of technology makes a crime easier to commit for TEHC.

With the adoption of technology for the perpetration of scams and the increasing evidence suggesting that scams are becoming technology driven (IC3, 2008, and ABS, 2007), cyberscams are overwhelmingly taking the place of more traditional-based scamming tactics (Choo et al., 2008, and Denman et al., 2004). Cyberscams are Internet assisted scams and fundamentally, cyberscams contain characteristics compliant to both TEAC and TEHC. Scams communicated through alternative and more traditional methods such as face to face interactions – similar to the fraud types defined in Chapter 1 (*D*) often also extend into the use of technology for communications and funds transfers. Regardless of this, the definition offered by the AIC only recognises Internet assisted scams as TEAC type scams (Choo et al., 2008).

Cyberscams, or Internet assisted scams have long been known as traditional crimes perpetrated through the use of computer technology (Choo et al., 2008, and Wahlert, 1998). An example of a traditional scam is the Ponzi scam. Similar to the Pyramid scam, this type of scam has been known to authorities since 1919 when Charles (Carlos) Ponzi pioneered the first recorded Ponzi scam with his investment company; The Security Exchange Company. Ponzi offered investors a minimum 50% return on reserves within the first 90 days of investment (USH, 2009). Late 2008, Bernard Madoff was arrested with multiple charges of fraud and criminal offences. Through his NASDAQ stock exchange dealings (Creswell and Thomas, 2009), Madoff defrauded his clients by an estimated $US50 billion in a Ponzi-style investment scam. A traditional scam operating successfully (for a time) eighty-plus years after its debut. Even though these cases and others like them operate successfully with and without the use of computer technology, those perpetrated through the use of such are recognized as TEAC; where the target of the crime is the end-users machine, rather than TEHC; where the target of the scam is the end user which is concerning since a Ponzi scam clearly targets the person and not their computer. The description separating TEAC from TEHC leads to confusion over the correct categorisation of scams in general.

For the AIC, technology based crimes are defined by two criteria. The first criteria focuses on the initial target of the scam and this could either be the user (semantic) or machine (syntactic). The second criterion is the level of technological dependence; enhanced or enabled. The very nature of a syntactic crime suggests that technology is required during the inception, creation and dissemination of the scam. Since technology is necessary for the commission of a syntactic crime, these crimes clearly belong to TEAC. For the purposes of this research, the relationship between the 'target' and level of 'technology dependence' is demonstrated in Figure 10.

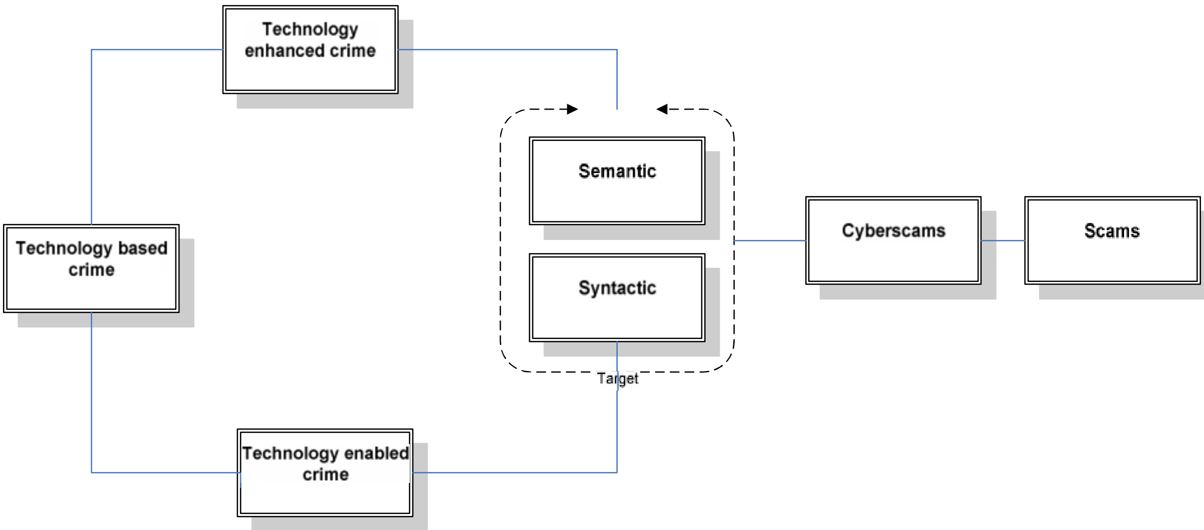


Figure 10: Redefined Technology Based Crime Flowchart

In this re-arrangement, the distinction between the 'target' and the level of 'technology dependence' is still present, however, rather than the target of the scam featuring as the first criterion for scam separation as it does with the AIC definition, it is substituted with the level of 'technology dependence' to form the first criterion within all technology based crimes. The re-ordering of 'target' and level of 'technology dependence' is necessary to form a useful definition of scams and Internet assisted scams that will be helpful in the identification of all scams rather than just certain types of scams. By rearranging these criterions the mistake of syntactic scam exclusion from this and future research will be avoided.

2.3 Retrospection

During 1998 Glenn Wahlert from the Australian Federal Police (AFP) presented a paper at the Internet Crime Conference (ICC), which was hosted by the AIC in Melbourne, Victoria. Primary concerns surrounding the growing trend of reliance on technology for businesses, the finance sector as well as general private use were raised (Wahlert, 1998). The issues discussed were highly pronounced within then-current scam survey results which were derived from victim report data and sourced from around the globe. During this time, technology based crime was predominantly recognised as TEAC and these were such things as virus infection and hacking. Wahlert (1998) extended the concept of technology based crimes beyond tech-as-target crimes to include semantically driven tactics and identified the banking and finance sectors, counterfeiting, sexually related crimes, gambling, and tactical intelligence as future targets of cyber-engaged criminals. Wahlert (1998) recognised four themes supporting the development and dissemination of Internet assisted scams; anonymity, mass communicability, jurisdictional impedances, and cultural ambiguities, suggesting that these areas were in need of urgent attention. It was predicted that a surge in cyber-criminal behaviour would be imminent suggesting that technology based crime would thrive with an identified focus on scam based tactics (Wahlert, 1998).

Wahlert (1998) suggested that effective mechanisms for the identification and monitoring of technology based crime were necessary for controlling its exploitation, further to this, it was recommended that the development of transnational agencies authorised to operate cross-jurisdictionally were imperative to fight the phenomena. Inter-agency cooperation and transnational coordination were identified as fundamental to the development of successful transnational operations, following the development of a shared language with which to correctly identify and monitor cases of technology based crime. More than a decade has passed since these revelations were made and resistance towards a consistent and uniform scam lexicon is still apparent (ACPR and AUSTRAC, 2006, Lea et al., 2009, Hays and Prenzler et al., 2002, and Stabek et al., 2009).

Stabek et al. (2009) recognise the breadth in scam descriptions and through a critical analysis of reports produced by the ABS, IC3, and the ERG, a compelling case was made justifying the need for a Consistent Cyberscam Classification Framework (CCCF). The authors suggest that scammers operate with three primary goals; of information gathering, participation seeking and financial gain (Stabek et al., 2009). It was suggested that information based scams target the recipient's personal data and scams falling into this category were identity theft and phishing scams. Participation seeking scams were described by money transfer, laundering and re-shipper scams and financial gain scams were described as possessing 'hit-and-run' style tactics where the scammer and victim are involved in a once off transaction.

Using these three goals; information – participation - money, complex scams can be described. An example of a complex or multi-goal scam is the upfront fee for loan or credit card scam. In this scam, the receiver is required to provide all personal and financial details to the scammer as well as pay an upfront fee. Here, the scammer's goal of information gathering is realised while the goal of financial gain from the up front payment is achieved.

2.4 Scams in Research

Scams are a tool for exploiting individuals; they are used by the criminal element to trick unsuspecting people into giving up something of value, whether it is money, information, or their time. Scams are assisted by the use of the internet and computer technology and can be either syntactic or semantically driven. The appearance of all scam types equally in research literature is unprecedented and the most commonly researched scams are internet auction scams, phishing scams, spam scams, Nigerian 419 scams, and advance fee fraud scams.

2.4.1 Internet Auction Scams

An Internet auction scam comes in five forms; shill bidding, bid shielding, merchandise non-delivery, payment non-delivery, and product authenticity (Dolan, 2004). These can also occur at the delivery phase of the scam involving fake escrow services or non-existent courier services. For the following explanation, let: $a \rightarrow e$ = the type of internet auction scam, x = the perpetrator and y = the victim. These definitions are described further in Figure 11.

i.	a= shill bidding: x = seller, y = customer
ii.	b = bid shielding: x = customer, y = seller
iii.	c = merchandise non delivery: x = seller, y = customer
iv.	d = payment non delivery: x = customer, y = seller
v.	e = product authenticity: x = seller, y = customer

Figure 11: Internet Auction Scams Defined

In a shill bidding scam the victim is the customer and the scammer is the seller. It involves the cooperative efforts of a team of individuals or the use of falsified identities where the aim of the scam is to drive up the bidding prices of auction items. In bid shielding the roles of the victim and the perpetrator are reversed, the scammer is the customer, the victim is the seller. These scams involve the cooperation between two or more individuals, or the use of falsified identities. The buyer and an associate work together by outbidding each other for an auction item, the highest bidder (the associate) drops out of the auction at the last minute and the scammer claims the item at the lower price. For merchandise non-delivery, the victim is the customer and the scammer is the seller, the customer sends the required funds for the purchase of an auction item and the goods are never received. With payment non-delivery, the customer is the scammer and the seller is the victim. The seller forwards the sold goods and the customer receives the goods while revoking payment for the purchased items. In the case of product authenticity, the seller is the scammer and the customer is the victim. The seller misleadingly advertises their goods as something that they are not and the customer is led falsely to believe that they are bidding on a genuine item (Dolan, 2004).

Internet auction scams can be complex and difficult to analyse because the roles of the victim and the scammer are not constant for all cases. The complexity of internet auction scams also increases with the malleable nature of these scams and the ability of a scam to start as one type of scam and transform into another. An example of this complexity can be demonstrated with a scam that starts as a shill bidding scam where a buyer and an accomplice work together to drive up the selling price of an item and then when the item does not arrive to the purchaser, a merchandise non-delivery scam has occurred. If no accomplice has been used during the shill bidding component of the scam and a false identity has been created to assist in the scam, then other criminal acts have occurred and the scam is no longer a fraudulent case alone, rather, an identity crime which may also involve identity theft and identity fraud; where identity theft is the fraudulent attainment of another persons personal information and identity fraud is the fraudulent use of such personal information (ACPR and AUSTRAC, 2006).

Much of the research to date on internet auction scams involves detailed case study analysis of the victims of such scams (Dolan, 2004). Dolan (2004) analyses the demographics of internet auction scam victims through surveying known victims. The results suggest that over 60% of victims are Bachelor Degree educated or higher with almost 70% of victims being aged between 25 and 45 years old. Merchandise and payment non-delivery scams were the most common methods of fraudulent activity experienced by the survey respondents with 67.3% falling within this category and eBay was the most used auction platform at 73.5%. Dolan (2004) recommends that the introduction of a centralised and dedicated Internet auction fraud investigative team and reporting system is necessary to increase identification, monitoring and removal of fraudulent activities from Internet auction host Websites.

Due to the variant nature of internet auction scams, it is difficult to achieve a consistent and reliable real-time scam detection rate. Chau and Faloutsos (2005) suggest a method to increase the detection rate of auction fraudsters. The research involved the extraction of features from Internet auction seller histories to systematically identify and detect fraudulent cases. Up to 17 features were identified from 115 eBay user accounts which included 43 known fraudster accounts. Features were extracted from publicly available transaction histories and these features were such constructs as the dollar amounts of purchases, fluctuations in sell prices, and transaction frequencies. The dependent variable for each of three experimental runs was the number of features; 8, 16, and 17 respectively. For each experimental run, cross-fold evaluation involving decision tree analysis was performed and an 83% true-positive identification rate of known frauds was achieved from the 17-feature experimental run, 80% accuracy for the 16-feature experiment and 75% on the 8-feature experiment (Chau and Faloutsos, 2005).

Chau and Faloutsos (2005) recommend future research to focus upon the criminal management of fraudulent scams, and it is suggested that greater insight into the processes involved in the lifecycle of scams would have a positive impact on identifying, monitoring and intercepting scam transactions which would save victims and businesses greatly.

2.4.2 Phishing Scams

A phishing scam involves the misrepresentation of a trusted host to encourage recipients to divulge their personal or private information. The type of information usually sought is private details such as bank account numbers and passwords and personal details such as full name and date of birth. These scams are most often associated with, but not limited to financial institutions such as banks and credit unions. The scammers contact their victims through an email which usually closely resembles that of a financial institution, or other reputable industry or business. Whether or not the receiver of the phishing email is associated with the institution is of calculated consequence and little risk to the scammer since these scams are dispersed in such large quantities through unsolicited spam emails that the scammers are guaranteed to reach some genuine customers of the targeted institution. The phishing email will raise concern over the receivers account information and request verification of account details. It will be suggested that this can be achieved if the receiver clicks on the associated link contained within the email which will direct the receiver to a spoofed or falsified Website. Due to the speed of connectivity and low cost of scam dispersion, the scammers do not require all email receivers to respond to the communication for the scam to be a success. These scams often mark the collection phase of much larger scams and open a doorway to more complex scams such as identity fraud, impersonation, overdrawn accounts, money laundering , credit card applications, and falsified loans (Moore and Clayton, 2007).

Moore and Clayton (2007) investigate the usefulness of phishing site removal procedures in protecting banking customers from fraudulent banking Websites. Through their analysis of visitor frequencies and phishing Website take-down times, Moore and Clayton (2007) suggest that the length of time a phishing site is active corresponds with the state of the phishing attack. Data was collected from participating Web hosts which allowed access to their host-phishing statistics. Using a purpose developed testing system, 700 phishing Websites were analysed. It was found that on average, phishing sites remained operational for up to 57 hours and attracted responses from over 30 victims. An analysis of visitor frequencies and length of activation time confirmed that timely phishing Website removal could assist in combating phishing attacks (Moore and Clayton, 2007).

Weaver and Collins (2007) perform a capture-recapture method of analysis on Internet phishing activity and cluster phishing sites according to scam genres. The authors separate phishing activities from other scam events by rating each site on its likelihood of hosting a phishing scam. The authors suggest that since successful phishing campaigns require the public advertisement of the site, phishing scams are distinct from all other scams (Weaver and Collins, 2007). Phishing sites were clustered by the selected features; target, URL, and address, performing capture-recapture estimation to analyse the ratio of phishing content. Four types of phishing scams were identified; *isolated, persistent, bursty,* and *corrupt.* Isolated phishing scams were described as short lived and limited to a few scam genres; persistent phishing scams were also limited by genre type though these scams lasted longer than isolated phishing scams. Bursty phishing scams could have numerous scam genres, though brief in duration. Finally corrupt phishing scams were similar to bursty scams while experiencing longer life-cycles (Weaver and Collins, 2007). These experimental results demonstrated that only 40% of phishing-style scams were positively identified by their host Website and 60% remained undetected.

Jagatic et al. (2007) explored the social data mining context of phishing by harvesting publicly available personal and relationship based data by investigating the accounts on social networking sites. The authors explored the gullibility of individuals by applying a experimental phishing attack on a selected population for which they were able to obtain the most social and personal information. Comparisons were made between a control group and a socially engineered experimental group. The results establish that while 16% of participants in the control group responded to the phishing scam by supplying the sought information and 72% of those in the socially engineered group responded with their personal details (Jagatic et al., 2007). While the researchers received ethical approval and full support from their research institution to carry out this research, the reaction received by those involved afterwards suggests that a great weakness in personal security was exposed.

2.4.3 Spam Scams

A spam campaign involves random bulk email dispersion. The scammers might purchase the email addresses of known end-users or they might access the email list of a business or industry. Spam emails are synonymous with mass marketing scams, phishing scams, chain mail and syntactic attacks. The aim of a spam scam is to trick an end-user into responding to a communication.

Anderson et al. (2007) explore the business methods adopted by spammers by identifying the infrastructure of scam hosts. The authors developed a technique of mining spam scams in real time by clustering destination Websites. The real-time mining of spam scam is achieved through a technique called 'Spamscatter' before the clustering of destination Websites. From a sample of 36,390 unique URL's, 2,334 scams were discovered on 7,029 different machines. Spam scams were found to contain regular content such as watches, pharmacy, software, male enhancement, phishing, and Viagra (Anderson et al., 2007) and it is concluded that almost all spam campaigns expire in less than 3-days (99%), while most last less than 2-days (90%), and at least 50% are short lived, lasting no more than 12 hours.

Calais et al. (2008) propose a methodology for characterising similar spamming campaigns. They use frequent tree patterns and attribute association analysis to cluster spamming campaigns. The campaign goal and method were identified as target features along with language, layout and URL's. Minor scam differences were ignored with the goal of minimising the effect of spam obfuscation. The results of this research demonstrate that data mining techniques are useful in determining spammer behavioural patterns (Calais et al. 2008).

"ScamSlam: An Architecture for Learning the Criminal Relations behind Scam Spam" by Airoldi and Malin (2004) proposes the use of purpose built software to discover the origins and criminal cells involved in the dissemination and perpetration of scam prototypes. The program contains two distinct stages; filtering and clustering. Spam emails are first filtered using a Poisson classifier which identifies the likelihood that a message is a scam by assessing its probability status based upon the number of words within the message. A message is labelled a scam if the counts of words are greater than the probability of it being a scam than not being a scam. Airoldi and Malin (2004) assume that counts occur according to the Poisson distribution (1):

$$P(X_{mv} \mid W_m U_{ve}) = \frac{e^{-WmUve} (W_m U_{ve})^{Xrm}}{X_{mv}!} \tag{1}$$

Where $X_{mv} = 0,1,2,..$ and $W_m > 0, U_{ve} > 0$

Where w_m is the length of the message in thousands of words, u_{ve} is the Poisson rate for unigram *v* in category *c* and X_{mv} denotes the counts for unigram *v* in message *m*. Next, the maximum likelihood of the estimates from the Poisson classifier were calculated using the maximum likelihood algorithm in (2):

$$U_{vc} = \frac{\sum m \in M_c^{xmv}}{\sum m \in M_c^{wm}} \quad \textit{for each } c \in C \tag{2}$$

And, r_m is the ratio used to determine whether or not a message is more likely to be a scam than it is not (3):

$$r_m = \frac{\Pi_{v \in V} P(X_{mv} \mid U_v Spam)}{\Pi_{v \in V} P(X_{mv} \mid U_V NoSpam)} \tag{3}$$

If it was found that r_m was greater than 1 for any message, that message would be classified as a spamming communication. Following the classification and identification stage, those scamming communications that were positively identified as spam were then clustered using un-supervised hierarchical cluster analysis and a single linkage method with the following distance measure (4):

$$dist(m\sim_i, m\sim_j) = \frac{\sum_{k=1}^{n} W_{ik} W_{jk}}{\sqrt{(\sum_{k=1}^{n} W_{ik}^2)(\sum_{k=1}^{n} W_{jk}^2)}} \tag{4}$$

A sample of 500 scam campaign-based emails was tested. The filtration phase of ScamSlam correctly identified 99% of all scam-based spam messages (Airoldi and Malin, 2004). From the clustering phase of ScamSlam it is learnt that a minimum of 50% of the tested spam scam messages could be traced back to 20 individuals or criminal cells (Airoldi and Malin, 2004).

2.4.4 Nigerian 419 Scams

In Nigeria, all cyberscams are pursuant under the Advance Fee Fraud and Other Offences Act of 2006 while scams and what is commonly referred to as Nigerian Fraud are pursuant under section 419 of the Nigerian Criminal Law Act (Dyrud, 2005, and Glickman, 2005). A 419 scam is most often received by spam email. It will detail a story of misfortune and seek to appeal to empathetic and opportunistic personalities (Lea et al, 2009). These scams promise great financial reward for ongoing monetary assistance and can start as a random email that the recipient was 'lucky' to receive or may begin as another scam such as a romance scam and turn into an ongoing emotional and financial drain for the victim. There are many guises and themes of Nigerian 419 Scams but the commonality is that the victim will receive nothing whilst losing everything. There are numerous terms used interchangeably when describing a 419 scam and these are, and not limited to; Nigerian letter, 419, Nigerian fraud, advance fee fraud, and West African fraud.

Dyrud (2005) performs case-study analyses on Nigerian 419 Scams and recognises a number of identifiable persuasive techniques used by scammers in their 419 scams. Over one hundred 419 email scams were analysed spanning across 10 months. Dyrud (2005) manually analysed each 419 email noting that the highest traffic of 419 communications was during the month of February (n = 20). The most prominent type of 419 scam was the investment style scam (n = 47 scams), followed by estate scams (n = 46 scams) and lottery-style scams (n = 18). '*Ad miseriocordium'* – an appeal for pity as well as trust are cited as primary psychological tools in the 419 scammer's toolbox (Dyrus, 2005). For a 419 scam involving *ad miseriocordium*, an email would be received from someone in an obvious role of authority. It would describe a horrific situation or event which would appeal to the receiver's sense of pity and empathy (Dyrud, 2005). A 419 scam designed to appeal to the receivers trust involved the misrepresentation of a situation in which the sender was relying on and 'trusting' in the receiver. In these situations, Dyrud (2005) reports that scammers were asking for trust from their victim by implying that they had bestowed their trust in them, implying that a reciprocal relationship should follow. The concept of psychological attribute analysis is strengthened by the UKOFT's research on The Psychology of Scams (Lea et al., 2009).

2.4.5 Advance Fee Fraud Scams

Advanced fee fraud scams require the recipient or applicant to pay a fee in advance for a service which is never received or is not what was described. These scams can operate alone or as a subsidiary component to more advanced scams. A qualitative analysis investigating over 400 advance fee fraud scams found that specific writing techniques were used to illicit a desirable response from the receiver (Holt and Graves, 2007). It was suggested that by developing an understanding and a methodology for examining the persuasive language employed in advance fee fraud scams, insight into the reasons behind victim responses may be learned. Using a grounded theoretical approach, Holt and Graves (2007) analysed advance fee fraud scam content identifying key static features through a manual analysis of each scam.

By categorising the scams based upon their derived static feature content, the authors identified 14 scam types; business solicitation, fixed fee transfer from bank, fixed fee transfer from barrister, over-drafted contract, charity message, lottery message, fixed fee from government, fixed fee from citizen, investment, banking transaction, fixed fee transfer to account, fixed fee transfer for investments, consignment, and fixed fee from diplomat (Holt and Graves, 2007). From their analysis, Holt and Graves (2007) concluded that advance fee fraud emails share characteristics of deceptive simplicity, which identifies with the receiver, and use of unique phrases which would give the reader an impression of authenticity. They also suggest that scam templates may be recycled to suit future scam campaigns.

2.5 Summary

Scams have been identified in this chapter as belonging to both technology enabled and technology enhanced crime groups and the structure of the commonly used technology based crime model supported by the AIC has been reassessed to ensure that syntactically driven scams will not be excluded from future analysis. The issues relating to technology based scams have been identified as anonymity, mass communicability, jurisdictional impedances and cultural ambiguities and it has been recommended that the development of transnational agencies authorised to operate cross-jurisdictionally are necessary in combating these issues. Further to this, it has also been acknowledged that inconsistencies in scam descriptions need to be standardised.

Five areas of scamming research have been investigated from internet auction scams, phishing scams, spam scams, Nigerian 419 scams to advance fee fraud scams. While most scamming research investigates the areas of phishing and spam, only a snap shot of this research is provided here. Five different internet auctions scams have been recognised and it has been suggested that the introduction of a centralised reporting system would assist in the identification, tracking and interception of scams (Dolan, 2004). It has been reported that feature analysis can assist in achieving 80% accuracy in identifying fraudulent Internet auction scams (Chau and Faloutsos, 2005).

It was discovered from using publicly available frequency data that phishing scams remain active for up to 3 days (Moore and Clayton, 2007). Through clustering phishing scams by their scam genres it was demonstrated that by identifying the type of phishing scam a scam belongs to from the four possibilities; isolated, persistent, bursty, and corrupt, greater accuracy could be achieved in host site identification of phishing scams (Weaver and Collins, 2007). Also using publically available data was Jagatic et al. (2007) who performed a social engineering experiment by researching their selected sample and targeting those whom provided the most information about themselves on public Internet forums with an experimental phishing campaign.

Those spamming scam campaigns presented here have been investigated by clustering techniques to identify business models (Anderson et al. (2007), pattern tree analysis has been used to identify characteristics of spamming campaigns (Calais et a. 2008), and hierarchical cluster analysis has been used to identify naturally occurring partitions in scam types by grouping scams according to similarity (Airoldi and Malin, 2004). Through the use of hierarchical clustering it was established that scam spam messages could be traced back to 20 independent groups of origin (Airoldi and Malin, 2004).

Case study analysis has been used to analyse Nigerian 419 scams (Dyrus, 2005) with the results identifying psychological constraints which are exploited by scammers. Grounded theory has also been used to identify scam static features through a manual content analysis of Nigerian 419 scams (Holt and Graves, 2007). Advance fee fraud scams have been researched using a qualitative methodology which involved a manual content analysis of writing styles which identified scam static features (Holt and Graves, 2007). From the identified scam static features, the authors categorised scams and identified fourteen types of advance fee fraud scams. A comparative snapshot of the strengths and weaknesses of the approaches investigated here appear below in Table 6.

Table 6: Strengths V Weaknesses Table

Authors	Year	Focus	Method	Strengths	Weaknesses
Dolan	2004	Internet auction scams	Case study analysis	Detailed insight into victim accounts	Narrow focus, only surveys known victims
Chau & Faloutsos	2005	Internet auction scams	Feature extraction	Focus is on the scam, identification of key features, increased identification rate of scam accounts	Narrow focus, single method of feature extraction, little explanation of features
Moor & Clayton	2007	Phishing	Quantitative analysis of phishing site takedown times	Evidence for policy reform	Results only transferrable to the online banking sector
Jagatic et al	2007	Phishing	Experimental harvesting of personal information from social networking sites	Gives a good insight into current computer user behaviour	Sample limited to a particular cohort and may not be representative of the population
Anderson et al	2007	Spam	Real time data mining	Novel approach using image shingling	Does not shed any light onto the origin, content or makeup of spam
Calais et al	2008	Spam	Qualitative, tree pattern and attribute analysis	Insightful of the transaction flow of spam scams, strong methodology	Narrow focus, more depth could be explored and process aligned with business processes
Airoldi & Malin	2004	Spam	Cluster analysis	Strong methodology builds a case for use if HCA	Narrow focus
Dyrud	2005	Nigerian 419	Qualitative content analysis	Detailed insight into scamming communications	Short time span, small sample, single researcher bias
Holt & Graves	2007	Advance fee scams	Qualitative, grounded theory, static feature analysis	Builds a case for use of a grounded theoretical approach in the content analysis of bodies of text and the identification of static features	Narrow focus, small sample

This overview of scamming research demonstrates that the types of methodologies applied to the research of scams are varied and widely subjective in nature. Clustering and content analysis are identified as being favoured approaches amongst the research literature and it is these approaches that are explored further throughout this research. The following section outlines the goals for this research and details the research questions in terms of research problems and concludes with the contributions that this research will make to research literature as well as in practice.

Chapter 3: Methodology

3.1 Introduction

There are numerous approaches detailed in the literature on quantising qualitative data (Agresti, 2002, Aranganayagi and Thangavel, 2009, and Berkhin, 2002, Birzer and Craig-Moorland, 2008, Calais et al., 2008, Choi, 2008, Fernandez, 2004, Holt and Graves, 2007, and Jie et al., 2004). The approach applied in this study follows a mixed methodology. The data is collected and stored in qualitative form, and it is then analysed in a quantitative way resulting in quantitatively derived and empirically assessed conclusions. The approaches used in this study change throughout the varying stages of sampling, data identification, data collection, scam identification and scam group membership verification. Each method used within each data phase is outlined below.

3.1.1 Sampling

The data for this research was conveniently sampled from publicly available information sources consisting of Internet Websites and published reports. The scam descriptions from 14 sources belonging to 4 different countries was collected and those countries contributing data for this research were Australia, Canada, the United Kingdom and the United States of America. The data sources were manually inspected to verify their suitability for use of data collection. What determined source appropriateness was the presence and accessibility of text based scam descriptions. If an inspected data source did not contain written descriptions of scams listed on their Website or in their report, the source was discarded from further analysis. Due to the manual nature of source sampling, it was required that scam descriptions be written in English, and they also had be publicly available sources so that any member of the public could access them freely.

Table 7: Contributing Source Scam Frequencies

Source	Frequency
Scamwatch	40
Australian Competition and Consumer Commission	35
United States Postal Inspectors Service	33
Looks too good to be true	28
Scam smart	28
United Kingdom Office of Fair Trading	27
Internet Crime Complaint Center	10
Federal Bureau of Investigation	13
Environics Research Group	12
FIDO	12
On guard online	10
Australian Bureau of Statistics	8
US-Cert	7
Queensland Police Service	5

Of the sources listed above in Table 7, all were found on the Internet through completing a general search in both Google and Yahoo search engines for topical scam words and phrases such as ‘scam’, ‘fraud’, ‘swindle’, ‘technology based crime’, ‘scammers’, and ‘Internet crime’. Scamwatch, United States Postal Inspectors Service (USPIS), Looks too good to be true (L2G2BT), Scamsmart, Internet Crime Complaint Centre (IC3), Federal Bureau of Investigation (FBI), FIDO, On guard online (OGO), US-Cert (USC), and the Queensland Police Service (QPOL) all contained Websites with the necessary information while the Australian Competition and Consumer Crime Commission (ACCC), UK Office of Fair Trading (UKOFT), Environics Research Group (ERG), and the Australian Bureau of Investigation (ABS) were text based documents available for download on the Web.

Once a data source was found identified as useful, those scams detailed within the source were printed along with their source-based categorisation, title and description. The collected scam descriptions were then manually analysed by content analysis. In total, the sample consisted of 277 individual scam case descriptions, gathered from 14 publicly available sources. Collectively, the data sources had categorised the 277 scams into 38 different scam types which were later analysed based upon 82 pre-identified scam static features.

3.1.2 Data Identification

Feature identification has been used in the past in the investigation of scamming events (Chau and Faloutsos, 2005 and Weaver and Collins, 2007), regardless of this, no comprehensive list of scam static features relevant to all types of scams is known. Without a comprehensive list of scam static features it is difficult to determine scam membership. A purpose developed list of divisively derived scam static features pertinent to all scam types was subsequently developed which was achieved using a content analysis, a qualitative, grounded theoretical approach similar to the one used by Holt and Graves (2007).

By using a grounded theoretical approach, each scam description was regarded in its complete form and all of the information gathered to compile each scam descriptions therefore had an impact on the derived scam static features. Through a rigorous bottom-up interrogation similar to the approach used by Lamp et al. (2007), the scam static features emerged. While this method is subjective to pre-identifiable bias; researcher, scam description author location, nationality, and jurisdiction as well as interpretational ambiguities, the data which would become the static features for this research was derived from evaluations of scam descriptions and it was guided by pre-defined source comparative features found in the form of 'what to watch out for' statements in the source data. Due to this, the 'what to watch out for' section of scam descriptions became a useful tool guiding to identification of scam static features. To explain this further, an example is given below in Figure 12 which was sampled from the Scamwatch Website.

Warning Signs

- You receive an email or a phone call from somebody saying they are from your bank, asking you about recent activity on your credit card or account.
- You are asked to confirm your credit card and bank account details by return email, visiting a website or over the phone.
- The caller or the email claims that there has been fraudulent activity found on your bank account, or that your card has been cancelled.
- You may be advised to contact a fake fraud investigations body, and discouraged from contacting your bank or credit union.
- www.scamwatch.gov.au

Figure 12: Example of 'What to Watch Out For" from the Scamwatch Website

Four key points are identified above advising the reader of things to watch out for if they wish to avoid becoming involved in a Phony Fraud Alert. The potential victim should be alarmed if they have *received* a communication by either *email* or *phone* from somebody *claiming* that they are from their *financial institution*. If they are *required* to provide their *banking details* because a *claim* is made that there has been *fraudulent activity* on their account and it is implied that the their card or account will be cancelled. They may then be *discouraged* to contact their financial institution and to *remain silent* about the issue.

From this example, some of the static features derived are the role of the victim – *receiver*, the method of scam contact – *email* or *phone*, something the scammer claims – *from a financial institution* and that there has been *fraudulent activity,* something the victim is required to do – *provide account details*, and scammer tactics to support the success of the scam – *discourage* the victim from seeking assistance, and require them to *remain silent* about the exchange.

Using the pre-identified key features from the data sources, a list of scam static features was compiled containing 82 individual features from 9 categories of feature-type. Once compiled, a scam static feature list was used during the examination of each scam description and the presence or absence of the identified feature within the descriptions was recorded in a progressive spreadsheet in binary form.

A comprehensive list of scam static features along with feature-type appears in Table 28 (in the Appendix). The role that a victim could play in a scam was either as a seller, a customer or a client, or random for un-associated attacks. The method of scam introduction was another feature type, the scam was either sought by the victim; ticketing scams are an example of this where the victim acts as customer seeking tickets for an event and unwittingly involves themselves in a scam by purchasing non existent tickets from a fake ticketing agent. Scams received by the victim can be described to lottery scams received in the post or by email which suggests that the receiver has won a substantial prize in a lottery or competition that they have not heard of nor entered. Or, scams that are introduced to the victim such as those face to face scams like door to door scams offering services that are never performed or pyramid style investment scams that are most often introduced to the victim by a friend or a known acquaintance.

The tool for scam proliferation category relates to the method of initial scam receipt, this could have been achieved in one or a combination of 11 ways from emails to pop-ups, Websites and Internet forums to face to face introduction, post, telephone and text messages. Seventeen different scam offers were identified by the data sources and these ranged from human interaction to prizes, money and merchandise. It was identified that claims were often made by scammers to involve the receiver in their scam, 12 scam claims were recognised and these ranged from the claim that the proposal was government approved and legal, or the venture being of little risk, high reward and effective.

For a scam to be a success, the scammer must acquire what he or she seeks from the receiver, for this purpose, the 'what the scam required from the victim' category was introduced and contains 14 different features ranging from supplying personal and banking details to recruitment of other people. The 'method of the scam' refers to the target of the scam; semantic or syntactic, and the 'scammers toolbox' contains the features that are regarded as scamming extras used by scammers to aid in the success of the scam. This feature type contains 14 static features from the use of a compromised or falsified Website to paraphernalia, testimonials and story telling strategies.

Finally, the primary target of the scam as reported by Stabek at al. (2009) contained three categories; information, money and participation. A scam targeting an individual for money and money alone such as a once off advance fee to claim a non existent prize would be a financial gain scam, where a scam requesting bank account detail updates might target the gathering of information, and a scam requiring the involvement of the receiver in any way such as a money transfer employment style scam would be a participation style scam. Some scams contain more than one objective such as and advance fee loan scam which requires the payment of an advance fee as well as the submission of personal details to the scammer, this sort of scam would be a financial gain and information gathering scam.

3.1.3 Data Collection

Once a comprehensive list of static features was compiled, the collected scam descriptions were then manually analysed using a content analytical approach and the absence and presence of scam features was recorded. A content analysis approach has been used by Dyrud (2005) and Holt and Graves (2007) and is identified a useful approach for investigating scam incidents by Jie et al. (2005). The process of identifying absent and present scam static features was arduous and time intensive. The static feature analysis of scam descriptions represented in the vector space contains 277 scam cases with 82 features for each scam case entry, an example of the vector space model appears below in Table 8 where the columns represent the features and the rows represent a single scam case.

Table 8: Example of Vector Space Model of Scam Cases and Static Features

Scams	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Phishing	0	1	1	0	1	1	1	0	0	1
Identity theft	1	1	1	0	0	1	1	1	1	0
Romance	1	0	1	1	1	0	0	0	1	0
Internet dialer	1	1	0	0	0	1	0	0	1	1
Clairvoyant	0	1	1	1	0	0	1	1	1	0

3.1.4 Scam Group Membership Identification

The interest of this research is in grouping like scams based upon their static features. To achieve this, cluster analysis is used to partition scams. Various clustering methods have been applied in the research of scams (Airoldi and Malin, 2004, Anderson et al., 2007, Calais et al., 2008, Chau and Faloutsos, 2005, Choi, 2008, Dolan, 2004, Dyrud, 2005, Glickman, 2005, Holt and Graves, 2007, Jagatic et al., 2007, Lea et al., 2009, Losch, 2006, Moore and Clayton, 2007, and Weaver and Collins, 2007). The method used for this analysis is hierarchical cluster analysis and is similar to that approach used by Airoldi and Malin (2004).

K-means cluster analysis (Abonyi et al., 2007) is one of the most common forms of cluster analysis and is represented by the following algorithm (5) which aims to cluster N data points into X clusters with the minimum sum of squares:

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} |x_n - \mu_j|^2, \qquad (5)$$

It is a partition-based form or cluster analysis that relies on cluster centres of a pre-specified number of sought after clusters. It measures the distance between clusters using the Minkowski metric (6):

$$d_p(x_i,x_j) = (\sum {}^{d}_{k=1} |x_{i,k}-x_{j,k}|\mathrm{p})^{1/p} = ||x_i-x_j||_{p} \qquad (6)$$

The Euclidean distance algorithm (7):

$$d_2(x_i,x_j) = (\sum {}^{d}_{k=1} (x_{i,k}-x_{j,k})^2)^{1/2} = ||x_i-x_j||_{2} \qquad (7)$$

Or the Mahalanobis distance (8) algorithm:

$$d_m(x_i,x_j) = (x_i-x_j)F^{-1}(x_i-x_j)T \qquad (8)$$

This method of cluster analysis is a partition-based method of cluster analysis unlike agglomerative hierarchical clustering which, as the name suggests, is an agglomerative method of cluster analysis. K-means analysis requires data of a continuous or quantitative nature, the researcher must also know how many cluster they wish to find. For the data being investigated here, there was no expectation for how many clusters that would be found and it is for these reasons that it was determined that k-means cluster analysis was not suitable for use on the binary scam data gathered for this investigation.

Using an unsupervised agglomerative hierarchical cluster approach, starting with all scam cases, like scam cases are partitioned into analogous groups, this process continues until all scam clusters merge to form one large and final cluster. Unsupervised agglomerative hierarchical clustering assists in grouping scams into homogeneous scam genres, since agglomerative clustering is limited by an inability to divide pre-grouped clusters, it was selected as the optimum method for clustering scam cases because a logical tie or combination could not be overrun by a future connection. Hierarchical clustering was also selected as a suitable method for finding homogeneous partitions for this data set due to the size of the sample collected and because it was unknown how many cluster centres would be found within the data. Due to this, other forms of cluster analysis were unsuitable. K-means cluster analysis relies on knowledge of how many cluster centres exist (Agresti, 2002, Berkhin, 2002, Francetic, 2005 and Witten and Frank, 2000) and the two-step cluster analysis method requires mixed variables – qualitative and quantitative data. The data for this research is binary and there is no expectation for the number of cluster centres, therefore, the exploratory results offered by the unsupervised agglomerative hierarchical cluster procedure, its suitability for smaller data sets, as well as its appropriateness for binary data meant that this method was the optimal choice for the clustering of scam cases.

There are 8 different binary distance measures suited for use with exclusively binary data. These are (Meyer et al., 2004) the Jaccard coefficient:

$$a/(a+b+c) \tag{9}$$

The Sorenson-Dice coefficient:

$$2a/(2a+b+c) \tag{10}$$

The Anderberg coefficient:

$$a/(a+2(b+c)) \tag{11}$$

The Ochiai coefficient:

$$a/(\sqrt{(a+b)(a+c)}) \tag{12}$$

The Simple Matching coefficient:

$$(a+d)/(a+b+c+d) \tag{13}$$

The Rogers Tanimoto coefficient:

$$(a+b)/(a+d+2(b+c)) \tag{14}$$

The Ochiai 2 coefficient:

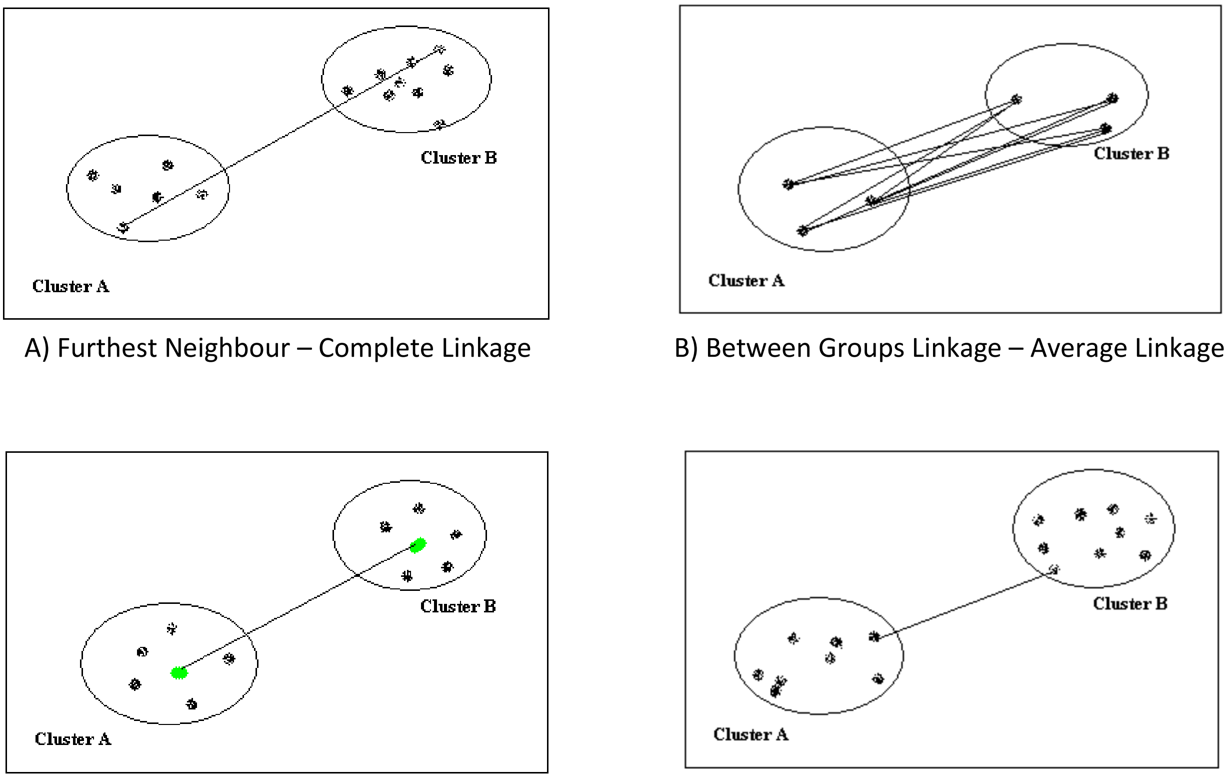
$$(ad)/\sqrt{(a+b)(a+c)(d+b)(d+c)} \tag{15}$$

The Russel Rao coefficient:

$$a/(a+b+c+d) \tag{16}$$

The two most suited binary distance coefficients to the data type used in this research were identified as the Jaccard coefficient and the Simple Matching coefficient; these were used for comparison to identify which was best for binary-vector based scam static feature-based data. The Simple Matching coefficient gives equal weighting to matches and non matches within the sample for binary data where 0 = absence and 1 = presence of a value while the Jaccard coefficient removes joint absences before giving equal weighting to both matches and non matches (Berkhin, 2002). While both distance coefficients are suited to binary data, the difference in the way that the zero response variables are accounted for might impact on the scam group placement and therefore, both of these distance measures are used and the results compared. A comparative study (Meyer et al., 2004) analysed the results of eight binary similarity coefficients; Jaccard, Sorensen-Dice, Anderberg and Ochiai, Simple Matching, Rogers and Tanimoto, Ochiai 2, and Russel and Rao. The results conclude that similar principles guide each different measure. A distinction was made between the Jaccard, Sorensen-Dice, Anderberg and Ochiai with the Simple Matching, Rogers and Tanimoto, and Ochiai 2 coefficients. The first four methods; Jaccard, Sorensen-Dice, Anderberg and Ochiai ignored zero value co-occurrences while the following three; Simple Matching, Rogers and Tanimoto, and Ochiai 2 incorporate these into their results. The final coefficient, the Russel and Rao is limited by the inclusion of co-occurrences in the denominator alone (Meyer et al., 2004).The Jaccard and the Simple Matching coefficients were selected for comparison because throughout statistical literature, these two methods are recommended more prominently for use with binary data (Alhaija and Richardson, 2003, Aranganayagi, Hennig, 2007 and Thangavel, 2009, and Berkhin, 2002).

For each binary distance coefficient, four linkage methods were tested. These were the furthest neighbour, between groups linkage, within groups linkage, and nearest neighbour coefficients. These methods were selected because of their suitability for clustering binary data. Furthest neighbour linkage is a method of complete linkage. It starts by grouping cases based on their maximum distance, linking all cases within clustered groups based on the furthest distance cases. Between groups linkage is a method of average linkage which starts by initially grouping the first set of cases by the maximum distance then clustering distances are created by calculating the average distance between all clustered cases. For within groups linkage, also known as centroid linkage, cases are grouped around the centre, or middle of a formation of clusters. Finally, for nearest neighbour linkage, also called single linkage, the shortest path between two cases is joined to create the first cluster and then continues in this manner until all cases are joined (Berkhin, 2002), these different linkage methods are presented figuratively below in Figure 13 and more descriptively in Table 9.



A) Furthest Neighbour – Complete Linkage

B) Between Groups Linkage – Average Linkage

C) Within Groups Linkage – Centroid Linkage

D) Nearest Neighbour Linkage – Single Linkage

Figure 13: Cluster Linkage Methods

Table 9: Linkage and Distance Measures

Linkage	Description	Jaccard	Simple matching
Nearest neighbour	$dist=b_{cl}-a_{cl}$ distance between the two closest objects from clusters *a* and *b*	$a/(a+b+c)$, [0,1]	$(a+d)/(a+b+c+d)$, [0,1]
Between groups	$dist=sum_{a,b1\ldots k}(b_{far1}-a_{far2})/n$ average distance between all objects from clusters *a* and *b*	$a/(a+b+c)$, [0,1]	$(a+d)/(a+b+c+d)$, [0,1]
Within groups	$dist=c_{clb}-c_{cla}$ distance between the centroids of all objects in clusters *a* and *b*	$a/(a+b+c)$, [0,1]	$(a+d)/(a+b+c+d)$, [0,1]
Furthest neighbour	$dist=b_{far}-a_{far}$ distance between the two furthest objects from clusters *a* and *b*	$a/(a+b+c)$,[0,1]	$(a+d)/(a+b+c+d)$, [0,1]

3.1.5 Scam Group Membership Verification

To assess the results from the hierarchical cluster analysis, a multivariate approach is used through the use of a discriminant function analysis which is used to determine the suitability of the model created. Discriminant function analysis has been used in the past by Mauldin (2008), Birzer and Craig-Moreland (2008) and Pyryt (2004) to predict group membership and ascertain the reliability of predictive models. Discriminant function analysis is suited to categorical data and assists in identifying which features are most important to the model. In this case the cluster membership results from each hierarchical cluster analysis are the dependent variables while the static features are the independent variables. Birzer and Craig-Moreland (2008) use discriminant function analysis in policing research to separate interrelating variables and to determine and predict future models of membership.

3.2 Limitations

The methods selected in this research for data analysis were chosen because they had been successfully used in the past on either similar data types or in a related field to that being investigated here. One of the biggest limitations to this research is the subjectiveness and interpretability of the data during the data identification and data gathering phases. Since this research was manually sourced, coded and collected, confidence can be gained in a single subjective and interpretive view which was stable across the whole data identification and data collection phase. However, this manual process proved time consuming and limiting because the researcher was limited to English only data sources and bound by time. Given more time and assistance from non-English speaking individuals, the scope of the data sources during the data collection phase could be expanded to include a larger and more representative and comprehensive sample pertaining to cultural groups and language specificities could be attained. Due to the nature of the data sources belonging to similar, related or at times the same jurisdictions and countries, there is a possibility that scam types will be repeated across source platforms. Since scam descriptions are assumed to be authored by the source agency, this has not become an issue in the consideration of data suitability because the purpose of this study is to analyse and compare scam types across related jurisdictions. A key challenge for future research is to identify ways to automate the data identification and coding processes.

To the knowledge of this researcher, the combination of analyses used in this research on the type of scam static features derived from those publicly available descriptions has not been attempted before. Therefore, this study represents an exploratory study into the usefulness of scam static features in predicting group membership and identifying scam genres. Exploratory analyses are troubled by concerns with validity, reliability and reproducibility which are the reasons for testing the reliability of the hierarchical clustering of scam static features using a discriminant function analysis.

This chapter defines the methodological approaches applied to this body of research and justifies the selection of the selected methods of analysis; the following chapter, Chapter 6 details the analysis results. The following chapter details the results from each phase of analysis and concludes with the most suitable model for identifying scam clusters as well as reporting on those static features which significantly impact the placement of scam cases into scam genres.

Chapter 4: Results

4.1 Data Summary

A total of 277 scam descriptions with 82 static features were analysed originating from four different English speaking countries; Australia, United States of America, United Kingdom, and Canada. Of the total number of scams recorded, 46.2% were from Australia, 39.7% from the USA, 9.7% from the UK and 4.3% were from Canada. The agencies from which the scams came and the number of scams attributed to their source appear in Table 10 below.

Table 10: Scam Frequencies by Source

Source	Frequency
Scamwatch	40
Australian Competition and Consumer Commission	35
United States Postal Inspectors Service	33
Looks too good to be true	28
Scam smart	28
United Kingdom Office of Fair Trading	27
Internet Crime Complaint Center	10
Federal Bureau of Investigation	13
Environics Research Group	12
FIDO	12
On guard online	10
Australian Bureau of Statistics	8
US-Cert	7
Queensland Police Service	5

Scamwatch provides most of the scams for the sample (14.4%) followed by the Australian Competition and Consumer Commission (ACCC) (12.6%). The 38 source-classified categories along with scam frequency appear in Table 11 below. The scams in the top ten scam categories make up nearly 73% (73.84) of the sample. The most frequent classification is Internet (n = 35) followed by

Mass Marketing (n = 26). There were 8 categories that contained no more than 3 scam cases, combined, these account for up to 10% of the sample. Identity fraud and identity theft were separated into two separate categories, suggesting that different types of identity focused scams exist, combined they make up over 5% of the sample.

Table 11: Scam Category Frequencies

	Category	No.	%		Category	No.	%
1	Internet	35	14.77	20	Identity theft	4	1.69
2	Mass marketing	26	10.97	21	Health insurance	4	1.69
3	Financial fraud	19	8.02	22	Chain letter & pyramid	4	1.69
4	No classification	18	7.59	23	Affinity fraud	4	1.69
5	Investment	18	7.59	24	Dating & romance	3	1.27
6	Lottery / competition / sweepstakes	14	5.91	25	Career opportunity scams	3	1.27
7	Job & employment	14	5.91	26	Telemarketing	2	0.84
8	Email scam	12	5.06	27	Psychic & clairvoyant	2	0.84
9	Advance fee scheme	10	4.22	28	Old fashioned fraud schemes	2	0.84
10	Auction fraud	9	3.80	29	Counterfeit payments	2	0.84
11	Money transfer requests	8	3.38	30	Charity scams	2	0.84
12	Miscellaneous	8	3.38	31	Telephone investment fraud	1	0.42
13	Identity fraud	8	3.38	32	Social engineering	1	0.42
14	Small business	7	2.95	33	Pharmacy	1	0.42
15	Mobile phone	7	2.95	34	Nigerian 419	1	0.42
16	Banking & online accounts	7	2.95	35	Door to door scams	1	0.42
17	Scam	6	2.53	36	Business fraud - opportunities	1	0.42
18	Health & medical	6	2.53	37	Betting & computer software	1	0.42
19	Fees for free Government services	5	2.11	38	Assassination / extortion	1	0.42

Table 12 below gives the frequencies of scam categories by country. All of the scams collected from Canada were categorised as mass marketing scams (n = 12) and of those scams sourced from the United Kingdom, 10 were not categorised, while 14 were identified as mass marketing scams, and 3 were career opportunity scams. The United States had scams classified into 19 different categories. Internet scams and financial fraud scams contained equal frequencies (n = 19) followed by Email scams (n = 10), and for this sample, four of those scams analysed received no classification at all. There were a total of 26 scam categories analysed from Australia and of the 26 categories, Internet scams contained the greatest frequency (n = 16) followed by investment scams (n = 14) and money transfer requests (n = 8).

From looking at these figures, conclusions can be drawn about scam emphasis in the source countries. Australia and the United States identify Internet scams as a type of scam while the United Kingdom and Canada do not coincide with this categorisation. Similarly, mass marketing scams are recognised by the United Kingdom and Canada as a scam category but not by Australia or the United States. Financial fraud appears to be a concern in the United States and small business scams, mobile phone scams, money transfer requests, banking and online account scams, scams, health and medical scams, identity theft, chain letter and pyramid scams, affinity fraud, dating and romance scams, psychic and clairvoyant scams, charity scams, telephone investment scams, door to door sales fraud, betting and computer software scams, and assassination or extortion scams are targeted scams within Australia.

Table 12: Scam Category Frequencies by Country

Category	Australia	United States	United Kingdom	Canada	No.	%
Internet	16	19	0	0	35	12.64
Mass marketing	0	0	14	12	26	9.39
Financial fraud	0	19	0	0	19	6.86
No classification	4	4	10	0	18	6.50
Investment	14	4	0	0	18	6.50
Lottery / competition / sweepstakes	6	8	0	0	14	5.05
Job & employment	8	6	0	0	14	5.05
Email scam	2	10	0	0	12	4.33
Advance fee scheme	9	1	0	0	10	3.61
Auction fraud	0	9	0	0	9	3.25
Money transfer requests	8	0	0	0	8	2.89
Miscellaneous	2	6	0	0	8	2.89
Identity fraud	2	6	0	0	8	2.89
Small business	7	0	0	0	7	2.53
Mobile phone	7	0	0	0	7	2.53
Banking & online accounts	7	0	0	0	7	2.53
Scam	6	0	0	0	6	2.17
Health & medical	6	0	0	0	6	2.17
Fees for free Government services	0	5	0	0	5	1.81
Identity theft	4	0	0	0	4	1.44
Health insurance	0	4	0	0	4	1.44
Chain letter & pyramid	4	0	0	0	4	1.44
Affinity fraud	4	0	0	0	4	1.44
Dating & romance	3	0	0	0	3	1.08
Career opportunity scams	0	0	3	0	3	1.08
Telemarketing	0	2	0	0	2	0.72
Psychic & clairvoyant	2	0	0	0	2	0.72
Old fashioned fraud schemes	0	2	0	0	2	0.72
Counterfeit payments	0	2	0	0	2	0.72
Charity scams	2	0	0	0	2	0.72
Telephone investment fraud	1	0	0	0	1	0.36
Social engineering	0	1	0	0	1	0.36
Pharmacy	0	1	0	0	1	0.36
Nigerian 419	0	1	0	0	1	0.36
Door to door scams	1	0	0	0	1	0.36
Business fraud - opportunities	1	0	0	0	1	0.36
Betting & computer software	1	0	0	0	1	0.36
Assassination / extortion	1	0	0	0	1	0.36

4.2 Feature Summary

Scam static features were categorised into nine different feature themes (Table 13). The role that the victim might play in a scam contained 4 features, the method of scam introduction contained 3 features, the method of scam dissemination was made up of 11 features, what the scam offered to the recipient contained 17 features, what the scam claimed it could do or offer the victim was made up of 15 features, what the scam required of its victims contained 13 features, the target of the scam contained 2 options – syntactic or semantic, the tools or tactics that a scammer incorporated into their scam contained 14 features, and the goal of the scammer could be one of, or a combination of 3 options see Table 9 below where 1 = the role of the victim, 2 = the method of introduction, 3 =the method of dissemination, 4 = what was offered, 5 = what was claimed, 6 = what was required, 7 = the target, 8 = the scammer's tools, and 9 = the goal of the scam.

While the identification of scam static features within scam descriptions was a bottom-up, grounded-theoretical process, the features identified here were assisted by those key points raised by each data source in the 'what to look out for' section of scam and scam category descriptions. It is entirely possible that static scam features are not limited to, or defined by the feature scams above. The importance of a feature for predicting scam membership of scam cases is explored further during the discriminant function analysis phase of this research.

To demonstrate how a scam can be defined by these static features an example of an internet auction scam is used. With an Internet auction scam, the victim might seek the scam out by acting as a customer, the victim therefore searches out the scammer. In this situation the victim who sought the scam through a website or online auction would be offered merchandise by the scammer. The scam may have claimed to offer the services of a refund if the victim was dissatisfied with their purchase, and to process the transaction, the victim may be required to pay for their purchase upfront and may have also been required to pursue alternative shipment methods to those ordinarily used by the online auction service. The target of this type of scam is the person not the machine and the tools that a scammer could use for this sort of scam might be the use of inferior merchandise or not providing the goods at all. A scammer's motivation for developing this type of scam is financial gain rather than information seeking or victim participation.

Table 13: Summary Table of Scam Static Features

Feature	Theme	Feature	Theme
Seller	1	Large return	5
Customer	1	Effective	5
Target Specific	1	Refund available	5
Un-associated	1	Fraudulent activity	5
Received	2	Share tips	5
Introduced	2	No credit check required	5
Sought	2	Little or no risk	5
Website / online auction	3	From corporate or government official	5
Face to face	3	Quick response	6
Text	3	Confidentiality	6
Phone	3	Pay up front costs	6
Seminar	3	Receive and send funds	6
Internet forum	3	Call premium number	6
Internet pop up	3	Transfer excess	6
Email	3	Complete sale outside of auction	6
Post	3	Send on to others	6
Advertisement	3	Recruit others	6
Fax	3	Supply personal information	6
Prize or money	4	Supply bank account details	6
Human Interaction	4	Invest	6
Financial return	4	Make a donation	6
Membership	4	Alternative shipment	6
Advice or assistance	4	Syntactic	7
Overpayment	4	Semantic	7
Treatment	4	Spoofed or fake website	8
Employment	4	Disguised as invoice	8
Opportunity for self or others	4	Inferior merchandise	8
Holiday	4	Falsified forms	8
Financial services	4	Paraphernalia	8
Good luck	4	Goods never sent	8
Property	4	Story based	8
Services	4	Veritiable street address	8
Merchandise	4	Looks genuine	8
Partial payment	4	Exploit of legitimate business	8
Insight	4	Testimonials	8
Legal	5	Reward greater than upfront cost	8
From financial institution	5	Further contact by email of phone	8
Detail update or confirmation required	5	Polite broken English	8
Government approved	5	Financial gain	9
Love affection or connection	5	Information	9
Government agency	5	Participation	9

4.3 Model Analysis

Using a minimum of three cluster groups inferred from Stabek et al. (2009), and running a single agglomerative hierarchical cluster analysis on the data for exploration, an upper limit of twelve cluster genres was identified. The range of cluster membership frequencies gathered for each hierarchical model was nine and this was the range between twelve and three clusters. The purpose of this approach was to look for homogeneous sets by identifying which hierarchical cluster solution contained the least number of clusters with the fewest number of scam cases in each cluster. Two binary distance measures were tested across four linkage methods. It is hypothesised that a smaller number of scam clusters can be found than the publicly acknowledged 38 which were recorded during the data collection phase.

4.3.1 HCA: Furthest Neighbour – Jaccard Coefficient

Bar charts of the selected group number membership frequencies for the furthest neighbour Jaccard coefficient hierarchical cluster model appear in Figure 14 and the dendrogram for this model can be seen in Figure 22 which can be seen in further detail at http://www.icsl.com.au/capability/identity-theft/scams. The bar charts A and B display a negative trend which begins to normalise in chart C. In chart E, the trend becomes more uniform and a strong negative trend is displayed from charts G through to J. For the 12-cluster solution, three groups contain 5 scams (clusters one, eight, and ten). The mode of the distribution is cluster three which contains 51 scam cases, the range of this 12-cluster solution is 46.

The 11-cluster solution displays similar trends to that of the 12-cluster solution. Clusters one and eight contain the least number of scam descriptions, 5 and 6 respectively. The modal cluster is the same for the 11-cluster solution as it is for the 12-cluster solution, cluster number three (n = 51) and the range is unchanged (n = 46). The 10-cluster solution shows a similarly shaped distribution to that of the previous 11 and 12-cluster solutions. It has one minimum cluster which contains 5 scam cases (cluster eight) and the mode is cluster three containing 51 scam cases, the range remains unchanged (n = 46).

A 9-cluster solution reveals a distribution that is beginning to normalise. The group containing the minimum number of scam cases has changed from cluster number eight which was stable across cluster solutions ten, eleven, and twelve to cluster six (n = 7). The mode of the distribution is 50 scam cases and this occurs in cluster three. The stability of cluster three is evident throughout the results since cluster three has remained the modal cluster for each cluster solution - nine, ten, eleven, and twelve. The range has changed from 46 scam cases in the previous cluster solutions to 43 in this 9-cluster solution.

The 8-cluster solution reveals a stronger negative trend than the previous nine, ten, eleven, and twelve cluster solutions. The group with the least number of scam cases is cluster number five (n = 18) followed by cluster one (n = 30). The modal cluster has changed from being consistently cluster three to cluster two (n = 73). The range has grown from 43 to 55 in this 8-cluster solution model. In the 7-cluster solution, the cluster memberships begin to even out becoming more uniform. Scam cluster four contains the least number of scam cases (n = 18), the cluster mode occurs at cluster number two (n = 73) and this is closely followed by cluster number one with 71 scam cases. The range is the same for the seven cluster solution as it is for the eight cluster solution (n = 55).

For the 6-cluster model, a negative distribution is displayed and there is no change in the modal scam clusters; two and one respectively. The scam cluster with the least number of scam cases is scam cluster four which remains unchanged from the seven cluster model. Where change is apparent is for scam cluster three which has grown from 31 scam cases in the seven cluster model to 52 scam cases in the six cluster model. The five cluster solution shows a negative distribution and some clear changes in scam group memberships. Scam cluster three has become the cluster with the least number of scam cases (n = 20) and the modal cluster has become cluster one which contains up to 148 scam cases, the range of this cluster solution is 128.

The 4-cluster solution reveals a strong negative trend. The scam cluster with the least number of scam cases is cluster number four (n = 20), the mode is scam cluster one (n = 160) and the range is 140. The results for the 4-cluster solution are similar to those found in the 3-cluster solution. A strong negative trend is evident, the cluster with the least number of scam cases is cluster number three with approximately 20 scam cases, the mode is cluster number one which remains stable at 160 cases and the range is still 140.

The cluster solutions with the most promising results are those with six, seven, eight, and nine clusters. This is because these cluster solutions show evidence of homogeneity as the scam cluster frequencies start to even out across the distribution. The results from cluster number five down to cluster number three do not provide a useful model that offers distinction between scam descriptions. This is because of the sheer size of cluster groups which can be recognized by comparing the ranges of the ten models which appear in Table 14.

The range for each cluster solution remains constant at 46 for cluster solutions ten, eleven, and twelve. It drops by three scam cases to 43 for cluster solution nine and then increase by 12 scam cases to a range of 55 for cluster solutions six, seven, and eight. At cluster number five, the scam range increases from 55 to 128. The scam clustering model which maintains the least amount of difference and therefore the lowest possible range is the model of most interest to this research; those cluster solutions are six, seven, eight and nine.

Table 14: Summary Table of Cluster Solutions for the Furthest Neighbour Jaccard Coefficient HCA Model

Cluster solution	Max	Min	Range
12	51	5	46
11	51	5	46
10	51	5	46
9	50	7	43
8	73	18	55
7	73	18	55
6	73	18	55
5	148	20	128
4	160	20	140
3	160	20	140

4.3.2 HCA: Between Groups Linkage – Jaccard Coefficient

Bar charts of the selected group number membership frequencies for the between groups linkage Jaccard coefficient hierarchical cluster model appear in Figure 15 and the dendrogram for this model can be seen in Figure 23. The 12-cluster solution reveals a negative distribution. Seven groups contain fewer than 5 scams each (one, three, seven, eight, ten, eleven, and twelve). The mode of the distribution is cluster number two containing 170 scam cases, and the range for this cluster solution is 165. The 11-cluster solution displays a similar distribution to that of the twelve cluster solution. Clusters one, three, seven, eight, ten and eleven contain the least number of scam descriptions, less than 5. The modal cluster is the same for the 11-cluster solution as it is for the t12-cluster solution, cluster two (n = 170) and the range is unchanged (n = 165). The 10-cluster solution shows a similarly shaped distribution to that of the previous 11 and 12-cluster solutions. It has five minimum frequency clusters containing less than 5 scam cases each (one, three, seven, nine, and ten) and the mode is still cluster two containing 170 scam cases, the range remains unchanged (n = 165).

A 9-cluster solution shows no change in distribution to that of the ten, eleven, and twelve cluster solutions. The clusters containing the minimum number of scam cases are clusters one, three, six, eight, and nine with less than 5 scam cases each. The mode of the distribution is 200 and this occurs in cluster two. The stability of cluster two is evident throughout the results since cluster two has remained the mode for each cluster solution; nine, ten, eleven, and twelve. The range has changed from 165 in the previous cluster solutions to 195 in this nine cluster solution. The 8-cluster solution displays a negative trend. The scam clusters with the least number of scam cases are cluster numbers one, three, six, and eight, containing no more than 5 scams. The modal cluster is still cluster number two with 200 scam cases and the range is still 195. In the 7-cluster solution, a negative distribution is evident. Scam clusters two, five and seven contain the least number of scam cases (n < 5) while cluster one the cluster with the greatest frequency (n = 210). The range has increased by 10 cases to 205.

For the 6-cluster model, a negative distribution is also displayed and there is no change in the modal scam cluster, clusters one with 210 scams descriptions. The clusters with the least number of scam cases are clusters two and five (n < 5). The 5-cluster solution shows a negative distribution and some clear changes in scam memberships. Scam cluster two has become the only cluster with the least number of scam cases (n < 5) and the modal cluster is still cluster one which contains up to 210 scam cases and the range of this cluster solution is 205. The 4-cluster solution reveals a very strong negative trend, similar to that seen throughout the other cluster groupings. The scam cluster with the least number of scam cases is cluster number four (n = 10), the mode is scam cluster one (n = 210) and the range has declined to 200. The results for the 3-cluster solution are similar to those found in the 4-cluster solution. A strong negative trend is evident and the cluster with the least number of scam cases is cluster number three with approximately 30 scam cases, the mode is cluster number one which remains stable at 220 cases and the range has reduced again from 200 to 190.

Table 15: Summary Table of Cluster Solutions for the Between Groups Linkage Jaccard Coefficient HCA Model

Cluster solution	Max	Min	Range
12	170	5	165
11	170	5	165
10	170	5	165
9	200	5	195
8	200	5	195
7	210	5	205
6	210	5	205
5	210	5	205
4	210	10	200
3	220	30	190

Table 15 displays a summary of the results of this model. After revising the cluster ranges, maximums and minimums, it is concluded that none of the models produced with the between groups linkage method and the Jaccard coefficient distance measure are useful in determining homogeneous groupings of scam memberships. All models from the twelve cluster model to the three cluster model contain a cluster which is made up of in excess of 150 scam cases and another cluster or multiple clusters with less than 5 scam cases in each. The ranges for each cluster solution are also large and for this reason, the between groups linkage, Jaccard coefficient hierarchical clustering method does not satisfactorily partition scam descriptions into homogeneous groups.

4.3.3 HCA: Within Groups Linkage – Jaccard Coefficient

Bar charts of the selected group number membership frequencies for the within groups linkage Jaccard coefficient hierarchical cluster model appear in Figure 15 and the dendrogram for this model can be seen in Figure 24. The 12-cluster solution displays a uniform distribution. Cluster number twelve contains the least number of scam cases (n = 10) followed by clusters three, eight, and nine (n = 13). The mode of the distribution is cluster number 7 containing 50 scams, followed by cluster two which contains 40 scam descriptions. The range of scam descriptions across clusters in this twelve cluster solution is 40. The 11-cluster solution displays a negative trend that is becoming more homogeneous with uniformity emerging after the mode towards the tail. Cluster eleven contains the fewest number of scam cases (n = 10) and clusters seven and nine follow with 12 scam descriptions per cluster. The modal cluster is no longer cluster seven as with the twelve cluster solution but is now cluster three with 65 scam cases. The range of this distribution is 55, fifteen more than the 12-cluster solution. The 10-cluster solution displays a similarly shaped distribution to that of the 11-cluster solution. Three clusters are equal in minimum frequency of scam cases which is 10 and the modal cluster is cluster three with 65 scam cases. Clusters one and two contain equal frequencies (n = 40) as do clusters five and seven (n = 25). The range remains unchanged (n = 55) to that of the 11-cluster solution.

The 9-cluster solution shows little change in distribution from that of the 10-cluster solution. The clusters containing the minimum number of scam cases are clusters six and nine with 12 scam cases. The mode of the distribution is 65 and this occurs in cluster three. The stability of cluster three is evident throughout the results thus far since cluster three has remained the mode for each cluster solution; nine, ten, and eleven. The range has not changed from the previous two cluster solutions (n = 55). The 8-cluster solution displays a negative trend however, there appears to be some evening out of scam cluster memberships suggesting an increase in homogeneity. The scam cluster with the least number of scam cases is cluster number eight containing 10 scam cases. The modal cluster is cluster number three with 75 scam descriptions, this is followed by cluster two with 73 cases and cluster one with 40 cases while the range is 63. Clusters five and six contain the same number of scam description cases (n = 25) as do clusters four and seven (n = 20). In the 7-cluster solution, a uniform distribution is emerging. Scam cluster seven contains the least number of scam cases (n = 13) and cluster two has greatest frequency with 130 scam memberships. The range has now increased to 117. Scam clusters four and five each contain 25 scam cases and cluster six has 20 while cluster three contains 18. Cluster one is the second largest cluster with 35 scam cases.

The 6-cluster solution shows a more negative trend than the 7-cluster solution. The cluster with the least number of scam cases is cluster six (n = 13) and cluster two is again the modal cluster with a total number of 130 scam cases. Clusters four and five are equal (n = 25) and the range of this cluster solution is the same as that of the seven cluster solution (n = 117). The5-cluster solution displays a strong negative distribution. Scam cluster five contains the least number of scam memberships (n = 13) and is followed by clusters three and four (n = 25). Cluster two contains the second highest scam memberships with 45 cases and the mode, which is cluster one contains 170 scam cases. The range of this cluster solution is 157 which is a 40 scam increase from both the 6 and 7-cluster models. In the 4-cluster solution the cluster with the least number of scam cases is cluster three (n = 25) followed by cluster four (n = 30) and then cluster two (n = 45). Cluster one contains the most number of scam cases (n = 170) which is the same as the previous five cluster solution and the range for this model is 145. The final hierarchical model using the within groups linkage method and the Jaccard distance coefficient is the 3-group cluster solution. The first cluster contains the highest frequency which is close to 200 scam cases (n = 195) while clusters two and three are relatively equal in their scam memberships with 40 and 35 scam cases respectively. The mode of this scam cluster distribution is 160 which is a 15 scam case increase from the previous four cluster model and 120 scam cases greater than the first 12-cluster model.

The cluster solutions for the within groups linkage Jaccard coefficient model appear in Table 16 the most promising models are those with eight and nine cluster solutions. This is because these cluster solution show evidence of homogeneity as the scam cluster memberships start to even out and these solutions are those which contain the fewest number of scam clusters with the least number of scam cases in each cluster. Those solutions from cluster number seven down to cluster number 3 do not provide a useful model which offers distinction between scam descriptions. This is because of the sheer size of cluster groups which can be recognized by comparing the modes of the ten models.

Table 16: Summary Table of Cluster Solutions for the Within Groups Linkage Jaccard Coefficient HCA Model

Cluster solution	Max	Min	Range
12	50	10	40
11	65	10	55
10	65	10	55
9	65	10	55
8	73	10	63
7	130	13	117
6	130	13	117
5	170	13	157
4	170	25	145
3	195	35	160

4.3.4 HCA: Nearest Neighbour – Jaccard Coefficient

Bar charts of the selected group number membership frequencies for the nearest neighbour Jaccard coefficient hierarchical cluster model appear in Figure 16 and the dendrogram for this model can be seen in Figure 24. A quick inspection reveals that in six of the cluster solutions; twelve, eleven, ten, nine, eight, and seven, there are two obvious clusters; cluster numbers one and seven for cluster models ten through twelve, cluster numbers one and six for cluster models eight through nine, and lastly cluster numbers one and five for the seven cluster solution. After this cluster solution, for all of the decreasing sequential cluster solutions six through to three, only one cluster is apparent. For all cluster solutions, cluster one contains the most number of scam cases and this hovers between 260 and 270 throughout each cluster solution.

After revising the cluster results presented in Table 17 it is concluded that none of the models produced with the nearest neighbour linkage method and the Jaccard coefficient distance measure are useful in determining homogeneous groups of scam memberships. All models from the twelve cluster model to the three cluster model contain a cluster which is made up of over 250 scam cases and either no other clusters or one other cluster containing less than 10 cases. The ranges for each cluster solution are also large and for this reason, the nearest neighbour linkage with Jaccard coefficient hierarchical clustering method does not satisfactorily partition scam descriptions into homogeneous groups.

Table 17: Summary Table of Cluster Solutions for the Nearest neighbour Jaccard Coefficient HCA Model

Cluster solution	Max	Min	Range
12	260	5	255
11	260	5	255
10	260	5	255
9	260	5	255
8	260	5	255
7	260	5	255
6	260	0	260
5	270	0	270
4	270	0	270
3	270	0	270

4.3.5 HCA: Furthest Neighbour – Simple Matching Coefficient

Bar charts of the selected group number membership frequencies for the furthest neighbour Simple Matching coefficient hierarchical cluster model appear in Figure 17 and the dendrogram for this model can be seen in Figure 25. The 12-cluster solution displays a negative distribution. Cluster eleven contains the least number of scam cases (n < 5) and cluster one contains the most (n = 105), the range of this distribution is 100. An 11-cluster solution maintains the negative distribution, the minimum cluster is cluster number ten (n < 5) and the maximum cluster is still cluster number one (n = 105), this cluster solution also has a range of 100. A 10-cluster solution has the same negative distribution as that of the 11 and 12-cluster models. The cluster with the least number of scam cases is cluster number nine (n < 5) and the maximum is again cluster number one (n = 105). The same range of 100 applies here for the 10-cluster solution as it did for the 11 and 12-cluster solutions.

Evening among cluster frequencies is apparent for the 9-cluster solution and homogeneity among groups is emerging. The cluster with the least number of scam cases is cluster number nine (n < 5) and the cluster with the greatest scam case frequency is cluster number one (n = 105). Clusters three, four, five, and seven are relatively equal with 30 cases in each cluster. The 8-cluster solution shows a negative distribution with a minimum cluster frequency of less than 5 for cluster eight and a maximum cluster frequency of 130 for cluster number one. The range of this distribution is 125 and clusters three, four, and five contain relatively equal cluster frequencies (n = 25). For the 7-cluster solution homogeneity is lost and a strong negative distribution emerges. Cluster number seven contains the least number of scam cases while cluster number one is the modal cluster containing 130 cases. The range of this distribution is similar to that of the eight cluster solution (n = 125).

The 6-cluster solution is negatively distributed with a minimum cluster frequency of less than 5 cases for cluster number six and a maximum cluster frequency of 130 cases for cluster number one, the range is 125. A 5-cluster solution continues similarly to the previous model with an strong negative trend. The cluster with the least number of scam description cases is cluster number five (n < 5) and the cluster with the most number of scam descriptive cases is cluster number one (n = 148). The range for this cluster solution is 143 scam cases. The 4-cluster solution closely resembles the result of the five cluster solution. It displays a strong negative distribution, cluster four contains the least number of scam cases (n < 5) and the first cluster contains the most number of scam cases (n = 148).

The range of this distribution is the same as the 5-cluster solution (n = 143). Lastly, the 3-cluster solution is negatively distributed, the cluster with the least number of scam cases is cluster number three (n < 5). Cluster two contains 100 scam cases and cluster one contains 175 scam cases. The range of this final distribution for the furthest neighbour method and Simple Matching coefficient is 170.

Table 18 below summarises the cluster results, it is concluded that none of the models produced with the furthest neighbour linkage method and the Simple Matching coefficient distance measure are useful in determining homogeneous groups of scam memberships. All models from the twelve cluster model to the three cluster model contain a cluster which is made up of in excess of 100 scam cases and a minimum cluster with less than 5 scam cases. The ranges for each cluster solution are also large and for this reason, the furthest neighbour linkage, Simple Matching coefficient hierarchical clustering method does not satisfactorily partition scam descriptions into homogeneous groups.

Table 18: Summary Table of Cluster Solutions for the Furthest Neighbour Simple Matching Coefficient HCA Model

Cluster solution	Max	Min	Range
12	105	5	100
11	105	5	100
10	105	5	100
9	105	5	100
8	130	5	125
7	130	5	125
6	130	5	125
5	148	5	143
4	148	5	143
3	175	5	170

4.3.6 HCA: Between Groups Linkage – Simple Matching Coefficient

Bar charts of the selected group number membership frequencies for the between groups linkage Simple Matching coefficient hierarchical cluster model appear in Figure 18 and the dendrogram for this model can be seen in Figure 26. The dendrogram displays decisive looking groups but the bar charts of cluster memberships display more useful information. A brief inspection of the bar charts reveals one cluster which contains almost all of the scam cases. In the 7-cluster solution, all of the 277 scam cases are lumped into one single cluster. This is evidence that the between groups linkage method and the Simple Matching coefficient cannot be used to partition scam description static features homogeneously across clusters and this is supported by the comparison of cluster solutions appearing in Table 19, therefore, this method cannot be used to satisfy the three research questions.

Table 19: Summary Table of Cluster Solutions for the Between Groups Linkage Simple Matching Coefficient HCA Model

Cluster solution	Max	Min	Range
12	210	5	205
11	210	5	205
10	210	5	205
9	225	5	220
8	230	5	225
7	270	5	265
6	275	2	273
5	275	2	273
4	275	2	273
3	275	2	273

4.3.7 HCA: Within Groups Linkage – Simple Matching Coefficient

Bar charts of the selected group number membership frequencies for the within groups linkage Simple Matching coefficient hierarchical cluster model appear in Figure 18 and the dendrogram for this model can be seen in Figure 27. Cluster 12-cluster solution displays a negative distribution, the cluster with the minimum number of scam cases is cluster number eight and it contains less than 5 scams. Cluster number two has the highest frequency (n = 102) and the range of this distribution is 97. The 11-cluster solution is similar to that of the twelve cluster solution. It has an almost identical distribution, the cluster with the least number of scam cases is cluster number seven (n < 5) and the cluster with the greatest number of scam cases is cluster number two (n = 120), the range of this distribution has increased to 115 from the twelve cluster solution. A 10-cluster solution reveals a similar negative trend to that of the eleven and twelve cluster solutions. The cluster with the least number of scam cases is still cluster number seven (n < 5) and the cluster with the most number of scam cases is cluster number two which remains stable with 120 scam case memberships. The range of this distribution is 115.

The 9-cluster solution remains almost unchanged from the 10-cluster solution since the distribution is still negatively skewed. The cluster with the least number of scam cases is cluster number seven (n < 5) and the cluster with the greatest scam case frequency is cluster number two which has grown from 120 to 135 scam cases. The range of this distribution has increased to 130 scam cases. The 8-cluster solution is negatively skewed, it contains a mode at cluster number two (n = 148) and minimum cluster at cluster number six (n < 5). The range of the eight cluster distribution is 143 scam cases. The 7-cluster solution displays a negative distribution, cluster two is the modal cluster (n = 152) and cluster number six is still the cluster with the minimum number of scam case memberships (n < 5). The range of the 7-cluster distribution is 147.

The 6-cluster solution contains a modal cluster at cluster number two (n = 152) and a minimum cluster at cluster number six (n < 5). Cluster number one contains 70 scam cases, cluster number three and cluster number four are relatively equal in scam cases memberships (n = 18) and the final cluster, cluster six contains 25 scam memberships. The range of this six cluster solution is the same as the range for the seven cluster solution (n = 147). For the 5-cluster solution, cluster number one still has a frequency of 70, cluster number two is still the modal cluster (n = 152), cluster number three and five contain 30 scam cases, and cluster number four is the cluster with the least number of scam cases (n < 5). For the 4-cluster solution, cluster number one contains 70 cases, cluster number two has 152 cases, and cluster numbers three and four are relatively equal in scam case frequencies with 30 scam cases each. The final cluster solution is the 3-cluster solution. The mode of this distribution is cluster number two and it contains 175 cases. Cluster number one still contains 70 cases and cluster number three is made up of 30 scam descriptions.

For each cluster model presented, a summary of the results appears in Table 20 there is one cluster which continually contains more than 100 scam cases. Due to this and the large ranges it has been concluded that this hierarchical cluster model is not suitable in determining homogeneous scam partitions using scam static features. This is evidence that the within groups linkage method and the Simple Matching coefficient cannot be used to partition scam description static features homogeneously across clusters and as with the between groups linkage model, this method cannot be used to satisfy the three research questions.

Table 20: Summary Table of Cluster Solutions for the Within Groups Linkage Simple Matching Coefficient HCA Model

Cluster solution	Max	Min	Range
12	102	5	97
11	120	5	115
10	120	5	115
9	135	5	130
8	148	5	143
7	152	5	147
6	152	5	147
5	152	5	147
4	152	30	122
3	175	30	145

4.3.8 HCA: Nearest Neighbour – Simple Matching Coefficient

Bar charts of the selected group number membership frequencies for the within groups linkage Simple Matching coefficient hierarchical cluster model appear in Figure 18 and the dendrogram for this model can be seen in Figure 28. Little investigation needs to be done to come to the conclusion that this model is unsuited to the homogeneous partitioning of scam cases. For every cluster solution only one cluster emerges and this cluster contains all of the 277 scam cases. Therefore, the nearest neighbour method and the Simple Matching coefficient cannot be used to partition scam description static features homogeneously across clusters and this method cannot be used to satisfy the three research questions, a comparison of the cluster results appears in Table 21.

Table 21: Summary Table of Cluster Solutions for the Nearest Neighbour Simple Matching Coefficient HCA Model

Cluster solution	Max	Min	Range
12	277	0	277
11	277	0	277
10	277	0	277
9	277	0	277
8	277	0	277
7	277	0	277
6	277	0	277
5	277	0	277
4	277	0	277
3	277	0	277

4.3.9 HCA Summary

The Jaccard distance coefficient produced some interesting results while the Simple Matching distance coefficient method did not partition adequately. A summary of these results appears below in Table 22. Using the furthest neighbour method, the Jaccard coefficient found four possible solutions and these were the six, seven, eight, and nine cluster models. The within groups method using the Jaccard coefficient also produced some interesting results with the eight and nine cluster solution. The purpose of the analysis was to identify homogeneous clusters of scam cases derived from scam static features. The hierarchical cluster model was determined to be suitable if it contained the fewest clusters with the least number of scam cases in each cluster. The above models have been selected for further analysis by discriminant function analysis to determine which model will best assist in predicting the accuracy of scam cluster memberships as well as identify those static features that are significant to the model.

Table 22: Summary of Cluster Solutions for Each HCA Method and Measure

Method / Measure	Jaccard	Simple Matching
Furthest neighbour	6,7,8,9	none
Between groups	none	none
Within groups	8,9	none
Nearest neighbour	none	none

It can be concluded by these results that the Simple Matching binary distance measure is not suitable for identifying homogeneous scam clusters inferred from derived scam static features. It can also be concluded that both the between groups and nearest neighbour linkage methods are unsuited to this type of data for clustering. The Jaccard binary distance measure and the furthest neighbour and within groups methods are the best models for determining homogeneous groups among scam cases derived from scam static features.

4.4 Model Verification

A discriminant function analysis is useful in developing rules which assign cases to clusters and because of this it has been selected to analyse the cluster memberships assigned to the scams during the hierarchical clustering analysis phase of this research. By using a discriminant function analysis, the predicted cluster memberships for each selected hierarchical clustering solution will be tested and conclusions drawn over the suitability of each model. The hierarchical cluster models selected for this procedure are the furthest neighbour method using the Jaccard distance coefficient for each of the six, seven, eight, and nine cluster solutions, as well as the within groups linkage method also using the Jaccard distance measure for an eight and nine cluster solution, where the cluster memberships formed due to the hierarchical clustering analyses are the dependent variables and the scam static features are the independent variables.

The goal of the discriminant function analysis is to assess the reliability of each selected hierarchical model. A reliable model is defined as a model in which the *(n-1)* discriminant functions combined account for at least 95% of variability within the data. A model which accounts for at least 95% of variability would then be concluded to be at least 95% accurate. Another goal of the discriminant function analysis is to identify which scam static features are significant to predicting scam cluster membership.

4.4.1 DFA: Furthest Neighbour – Jaccard Coefficient 9 Cluster Model

To predict the group memberships of the 9 clusters found in the furthest neighbour, Jaccard model, 82 predictor variables were used. First, the equality of group means was tested with the experimental hypothesis that group means are not equal. Of the 82 static features, 72 tested significant; these variables do not share mean equivalence across groups. There were nine variables that did not satisfy the experimental hypothesis and these were; good luck ($F = 1.043$, $p = 0.404$), insight ($F = 1.476$, $p = 0.166$), government agency ($F = 1.269$, $p = 0.260$), refund available ($F = 0.997$, $p = 0.439$), no credit check required ($F = 0.8$, $p = 0.603$), disguised as invoice ($F = 1.912$, $p = 0.058$), verifiable street address ($F = 0.547$, $p = 0.821$), further contact by email or phone ($F = 1.179$, $p = 0.312$) and polite broken English ($F = 1.820$, $p = 0.073$), a full table of results appears in Table 44 found in the Appendix. The variation within the model is not accounted for by any of these variables. One static feature was identified as failing the tolerance testing and this was the 'overpayment' variable.

From the Predicted Results table in Table 51 in the Appendix, the scam case number is given along with the cluster membership that was assigned to each scam case through the hierarchical clustering analysis along with its first and second predicted group memberships from the discriminant function analysis. From these results it can be seen that eight scams were not predicted accurately by the hierarchical clustering analysis. Scam case number 70 which is a scam recorded from the OFT and is titled Fake Clairvoyant Scam was clustered into cluster one by the hierarchical clustering analysis but its initial predicted group membership was scam cluster 3. The second cluster that this scam was predicted to most likely fit into was cluster number 4. Scam number 72 which was a Get Rich Quick Scam also sourced from the OFT was clustered into cluster 9 by the hierarchical clustering method but was predicted to belong to cluster 5 by the discriminant model. The alternative cluster which is predicted to most likely suit this scam is cluster 9; this is also the hierarchically clustered solution. Also assigned membership to its hierarchically clustered group during the second stage of prediction is the Psychic and Clairvoyant Mailing Scam (96), also by the OFT. This scam was assigned to cluster 1 during the hierarchical clustering procedure but was predicted to belong to cluster 3 by the discriminant function procedure.

The ACCC's Charity Scam (191) was assigned to cluster number 9 during the hierarchical clustering stage but was predicted to belong to cluster 2 by the discriminant function procedure. During the second stage of prediction however, this scam was allocated back to the 9th cluster. Multilevel Marketing (202) from the United States Postal Inspectors Service (USPIS) was clustered into scam genre 9 by the hierarchical clustering procedure but was predicted to belong to the 5th cluster during the discriminant analysis, and was reassigned back to the 9th cluster during the second stage of prediction. Unclaimed Tax Refund Scam (212) also from the USPIS was assigned to cluster 3 by the hierarchical clustering method but was predicted to belong to cluster 4, it was then reassigned back to cluster 3 during the second phase of predictions. Illegal Sweepstakes Information (220) by the USPIS was clustered into scam cluster 4 originally while it was predicted that it belonged to cluster 3, it was then reassigned to cluster 4 during the second predictive phase. Scam number 227 is Home Improvement and Repair Fraud sourced from the USPIS. It was clustered into scam cluster 4 by the hierarchical clustering procedure and predicted to belong to cluster 3 by the discriminant model. It was also reassigned to its original cluster during the second phase of prediction.

The Wilks' Lambda results in Table 50 in the Appendix confirm that each of the discriminant functions are significant to the model, all have p=values less than 0.05 ($p = 0.00$) and large Chi-square values. From the Eigenvalues table, Table 49 in the Appendix, it can be seen that the first function accounts for 32.8% of the total variation within the clustering of scam cases ($E = 14.997$). The first 6 functions capture 92.4% ($E = 2.353$) of the variation and by extending this to the 7th function, 96.3% ($E = 1.788$) of the variation in scam genre membership is accounted for.

4.4.2 DFA: Furthest Neighbour – Jaccard Coefficient 8 Cluster Model

To predict the group memberships of the 8 clusters found in the 8 cluster solution of the furthest neighbour, Jaccard model, the 82 predictor variables were used, the results of which appear in the Appendix. The equality of group means was tested with the experimental hypothesis that group means are not equal. Of the 82 static features, 68 tested significant; these variables do not share mean equivalence across groups. There were 13 variables that did not satisfy the experimental hypothesis and these were; fax ($F = 1.829$, $p = 0.082$), human interaction ($F = 1.367$, $p = 0.219$), good luck ($F = 1.064$, $p = 0.387$), insight ($F = 1.693$, $p = 0.111$), love affection or connection ($F = 1.647$, $p = 0.122$), government agency ($F = 1.137$, $p = 0.34$), refund available ($F = 0.63$, $p = 0.731$), no credit check required ($F = 0.601$, $p = 0.755$), from corporate of government official ($F = 1.749$, $p = 0.098$), send onto others ($F = 1.597$, $p = 0.136$), verifiable street address ($F = 0.627$, $p = 0.733$), further contact by email or phone ($F = 1.04$, $p = 0.403$), and polite broken English ($F = 0.671$, $p = 0.696$), a full table of results appears in Table 52 in the Appendix. The variation within the model is not accounted for by any of these variables. One static feature was identified as failing the tolerance testing and this was the 'overpayment' variable.

From the predicted group membership results in Table 56 in the Appendix, it can be seen that 7 scams were not predicted accurately by the hierarchical clustering analysis. Scam case number 70 which is a scam recorded from the OFT and is titled Fake Clairvoyant Scam was clustered into cluster 1 by the hierarchical clustering analysis but it's first predicted group membership was scam cluster 3. The second cluster that this scam was predicted to most likely fit into was cluster number 2. Scam number 72 which was a Get Rich Quick Scam also sourced from the OFT was clustered into group 8 by the hierarchical clustering method but was predicted to belong to group 4 by the discriminant model. The alternative cluster which is predicted to most likely suit this scam is cluster 8 which is the hierarchically clustered solution. Scam number 96, the Psychic and Clairvoyant Mailing Scam also by the OFT scam was assigned to cluster 1 during the hierarchical clustering procedure but was predicted to belong to cluster 3 by the discriminant function procedure. During the second stage of prediction however, this scam was reassigned to its original cluster, cluster 1.

Multilevel Marketing (202) from the USPIS was clustered into scam genre 8 by the hierarchical clustering procedure but was predicted to belong to the 4th cluster during the discriminant analysis, and was reassigned back to the 8th cluster during the second stage of prediction. Unclaimed Tax Refund Scam (212) also from the USPIS was assigned to cluster 3 by the hierarchical clustering method but predicted to belong to cluster 2; it then assigned to cluster 4 during the second phase of predictions. Scam number 227 is Home Improvement and Repair Fraud sourced from the USPIS. It was clustered into scam genre 2 by the hierarchical clustering procedure and predicted to belong to scam genre 3 by the discriminant model. It was reassigned to its original cluster during the second phase of prediction. The final scam that was not clustered by the hierarchical clustering procedure as the predictive model would suggest is the Cold Calling Scam from FIDO. It was clustered into scam genre 3 by the hierarchical clustering approach but it was predicted to belong to cluster 8 by the discriminant function analysis, however, its second group prediction reassigned Cold Calling back to its original cluster, cluster 3.

The Wilks' Lambda results in Table 55 found in the Appendix, confirm that each of the discriminant functions are significant to the model, have p=values less than 0.05 ($p = 0.00$) and large Chi-square values. From the Eigenvalues table, Table 54 in the Appendix, it can be seen that the first function accounts for 36.5% of the total variation within the clustering of scam ($E = 14.933$). The first 6 functions capture 95.7% ($E = 1.805$) of the variation.

4.4.3 DFA: Furthest Neighbour – Jaccard Coefficient 7 Cluster Model

To predict the group membership of the 7 clusters found in the 7 cluster solution of the furthest neighbour, Jaccard model, 82 predictor variables were used. The equality of group means was tested with the experimental hypothesis that group means are not equal. Of the 82 static features, 68 tested significant; these variables do not share mean equivalence across groups. There were 13 variables that did not satisfy the experimental hypothesis and these were; human interaction ($F = 1.601$, $p = 0.147$), good luck ($F = 1.234$, $p = 0.289$), insight ($F = 1.959$, $p = 0.072$), love affection or connection ($F = 1.928$, $p = 0.076$), government agency ($F = 0.883$, $p = 0.508$), refund available ($F = 0.557$, $p = 0.764$), no credit check required ($F = 0.649$, $p = 0.691$), from corporate or government official ($F = 2.048$, $p = 0.06$), send onto others ($F = 0.649$, $p = 0.691$), disguised as invoice ($F = 1.65$, $p = 0.134$), verifiable street address ($F = 0.469$, $p = 0.134$), further contact by email or phone ($F = 1.218$, $p = 0.297$), and polite broken English ($F = 0.732$, $p = 0.624$), a full table of results appears in Table 57 in the Appendix. The variation within the model is not accounted for by any of these variables. One static feature was identified as failing the tolerance testing and this was the 'overpayment' variable.

The scam case number is given along with the cluster membership that was assigned to it through the hierarchical clustering analysis and its first and second predicted group membership from the discriminant function analysis is seen in Table 61 in the Appendix. From these results it can be seen that 7 scams were not predicted accurately by the hierarchical clustering analysis. Scam case number 46 which is a scam recorded from IC3 and is titled Debt Elimination Scam was clustered into cluster 5 by the hierarchical clustering analysis but it's first predicted group membership was scam cluster 2. The second cluster most that this scam was predicted to most likely fit into was cluster number 5. Scam number 64 which was a Pyramid Scam sourced from the ABS was clustered into group 1 by the hierarchical clustering method but was predicted to belong to group 7 by the discriminant model. The alternative cluster which is predicted to most likely suit this scam is cluster 1 which is the hierarchically clustered solution. Scam number 72, Get Rich Quick by the OFT was assigned to cluster 7 during the hierarchical clustering procedure but was predicted to belong to cluster 3 from the discriminant function procedure. During the second stage of prediction however, this scam was assigned to its original cluster, 7.

Multilevel Marketing (202) from the USPIS was clustered into scam genre 7 by the hierarchical clustering procedure but was predicted to belong to the 3rd cluster during the discriminant analysis, and was reassigned to the 7th cluster during the second stage of prediction. Scam number 220, Illegal Sweepstakes Information from the USPIS was allocated within scam cluster 1 during the hierarchical clustering procedure and was assigned to scam cluster 2 by the discriminant function analysis. The second prediction for this case is its original cluster group, 1. Scam number 227 is Home Improvement and Repair Fraud sourced from the USPIS. It was clustered into scam genre 2 by the hierarchical clustering procedure and predicted to belong to scam genre 1 by the discriminant model. It was reassigned back to its original cluster during the second phase of prediction. The final scam that was not clustered as the predictive model would suggest is the Cold Calling Scam from FIDO (261). It was clustered into scam genre 1 by the hierarchical clustering approach but it was predicted to belong to cluster 7 by the discriminant function analysis. Its second group prediction however was the original cluster, cluster 1.

The Wilks' Lambda results in Table 60 in the Appendix confirm that each discriminant function is significant to the model, all have p-values are less than 0.05 (p = 0.00) with large Chi-square values. From the Eigenvalues table, Table 62 in the Appendix, it can be seen that the first function accounts for 31.6% of the total variation within the clustering of scam cases (E = 10.838). The first 5 functions capture 94.7% (E = 1.997) of the variation.

4.4.4 DFA: Furthest Neighbour – Jaccard Coefficient 6 Cluster Model

A discriminant function analysis of the 6 solution hierarchical clustering results reveals 14 static features that are not significant to the predictive model. These variables do not have a significant influence on the clustering results. These static features are human interaction (F = 1.916, p = 0.092), good luck (F = 1.487, p = 0.194), property (F = 1.773, p = 0.119), government approved (F = 1.440, p = 0.21), love affection or connection (F = 2.181, p = 0.057), government agency (F = 0.981, p = 0.43), refund available (F = 0.538, p = 0.747), share tips (F = 2.168, p = 0.058), no credit check required (F = 0.622, p = 0.683), send onto others (F = 2.252, p = 0.05), disguised as an invoice (F = 1.987, p = 0.081), verifiable street address (F = 0.565, p = 0.727), further contact by email or phone (F = 1.455, p = 0.205), and polite broken English (F = 0.721, p = 0.608) a full table of results appears in Table 62 in the Appendix. One final feature was removed from the model as with the three previous discriminant procedures and that is the 'overpayment' variable.

Table 66 in the Appendix details the results of the predicted groups memberships, only 6 scams were clustered differently to the predicted cluster memberships and all of these were replaced back into their original cluster group during the second phase of predictions. Debt Elimination from IC3 was scam number 46 and this case was originally placed into cluster number 5. The predicted cluster for this case is cluster 2 however. The Charity Scam (191) from the ACCC was the next case to be clustered differently. This case was originally clustered into cluster number 3 during the hierarchical clustering analysis while its predicted cluster membership is cluster number 2. Illegal Sweepstakes Information (220) from USPIS is the third case to be clustered differently. This case was placed into cluster 1 by the hierarchical procedure and was predicted to belong to cluster 2 by the discriminant function analysis. Scam number 227 is from the USPIS and its title is Home Improvement and Repair Fraud. It was originally placed into cluster 2 while its predicted cluster is cluster number 1. Scamsmart's Ponzi Scam (250) was placed into cluster 3 by the hierarchical clustering analysis and the discriminant function analysis predicted it to belong to cluster 1. Finally, FIDO's Cold Calling Scam (261) was placed into cluster 1 while it is predicted to belong to cluster 3.

The Wilks' Lambda results in Table 65 in the Appendix confirm that each of discriminant functions are significant, all have p-values less than 0.05 (p = 0.00) and large Chi-square values. From the Eigenvalues table, Table 64 in the Appendix, it can be seen that the first function accounts for 41.2% of the total variation within the clustering of scam cases. A four-function solution accounts for 92.9% of variation within the model.

4.4.5 DFA: Within Groups Linkage – Jaccard Coefficient 9 Cluster Model

A discriminant function analysis on the 9 solution hierarchical clustering result using within groups linkage and the Jaccard Coefficient reveals 20 static features that are not significant to the predictive model. These variables do not have a significant influence on the clustering results. These static features are text (F = 1.946, p = 0.063), fax (F = 0.969, p = 0.455), human interaction (F = 1.882, p = 0.073), holiday (F = 1.435, p = 0.191), financial services (F = 0.625, p = 0.735), good luck (F = 1.274, p = 0.263), property (F = 0.769, p = 0.614), services (F = 1.704, p = 0.108), insight (F = 1.841, p = 0.08), legal (F = 1.639, p = 0.125), government approved (F = 1.897, p = 0.07), love affection or connection (F = 0.791, p = 0.595), government agency (F = 1.956, p = 0.061), refund available (F = 0.459, p = 0.864), no credit check required (F = 1.342, p = 0.231), send onto others (F = 1.639, p = 0.124), verifiable street address (F = 0.598, p = 0.758), testimonials (F = 1.914, p = 0.068), reward greater than upfront cost (F = 1.451, p = 0.185), and polite broken English (F = 1.338, p = 0.233), a full table of results appears in Table 67 in the Appendix. One final feature was removed from the model as with 3 pervious discriminant procedures and that is 'overpayment'.

Table 71 in the Appendix details the predicted group memberships. Using this sample, 6 scams were clustered differently to the predicted cluster memberships. The first scam to have a different cluster prediction to its assigned cluster number is the Weight Loss Scam (12) from Scamwatch. It was placed into cluster number 6 from the hierarchical cluster analysis and its predicted cluster is cluster 2. The second prediction for cluster membership for this case is its original cluster solution 3. Scam 66, Credit or Bank Card Fraud from the ABS was originally placed into cluster 4, its predicted cluster is cluster 1 and its second predicted cluster is its original placement, 4. Scam number 112, Rolling Lab Scams from the FBI was grouped into cluster 6 by the hierarchical clustering procedure. Its predicted cluster membership was for cluster 1 while its second predicted membership is cluster 3. Romance Scams from 'Looks too good to be true' (137) was placed into cluster 7 but predicted to belong to cluster 3. This scam had a second cluster prediction that mirrored the hierarchical clustering results. Spam Scam (179) from the ACCC was grouped into cluster 7 while it was predicted to belong to cluster 3. The second predicted cluster solution for this scam was cluster number 4. Solicitations Disguised as Invoices (217) from USPIS was placed in cluster 8 and was predicted to belong to cluster 3. The second cluster that this scam was predicted to belong to was cluster 2.

The Wilks' Lambda results in Table 70 in the Appendix confirm that each of discriminant functions are significant, all have p-values less than 0.05 (p = 0.00) and large Chi-square values. From the Eigenvalues table in Table 69 in the Appendix, it can be seen that the first function accounts for 23.1% of the total variation within the clustering of scam cases (E = 13.899). The 7 function solution accounts for 95.4% (E = 2.789) of variation.

4.4.6 DFA: Within Groups Linkage – Jaccard Coefficient 8 Cluster Model

A discriminant function analysis on the 8 cluster hierarchical clustering result using within groups linkage and the Jaccard coefficient reveals 20 static features that are not significant to the predictive model. These variables do not have a significant influence on the clustering results. These static features are text (F = 1.695, p = 0.099), fax (F = 1.099m p = 0.364), human interaction (F = 1.741, p = 0.089), holiday (F = 1.578, p = 0.131), financial services (F = 0.878, p = 0.536), good luck (F = 1.406, p = 0.24), property (F = 0.670, p = 0.718), services (F = 1.517, p = 0.151), insight (F = 1.63, p = 0.116), legal (F = 1.924, p = 0.057), government approved (F = 1.661, p = 0.108), love affection or connection (F = 0.832, p = 0.575), government agency (F = 1.843, p = 0.069), disguised as an invoice (F = 1.798, p = 0.078), verifiable street address (F = 0.521, p = 0.84), reward greater than upfront cost (F = 1.287, p = 0.25), and polite broken English (F = 1.238, p = 0.277) a full table of statistics appears in Table 72 in the Appendix. One final feature was removed from the model and that is the 'overpayment' variable.

Tale 73 in the Appendix displays the predicted membership results. Using this sample, only 5 scams were clustered differently to the predicted cluster memberships. The first scam to have a different cluster prediction to its assigned cluster number is the Weight Loss Scam (12) from Scamwatch. It was placed into cluster number 7 from the hierarchical cluster analysis and its predicted cluster is cluster 2. The second prediction for cluster membership for this case is its original cluster solution 7. Scam 66, Credit or Bank Card Fraud from the ABS was originally placed into cluster 4, its predicted cluster is cluster 1 and its second predicted cluster is its original placement, 4. Scam 137, Romance Scam from 'Looks too good to be true' was placed in cluster 8 while its predicted cluster is cluster 3. This scam also had a second cluster prediction that mirrored the hierarchical clustering results. Spam Scam (179) from the ACCC was grouped into cluster 8 while it was predicted to belong to cluster 3. The second predicted cluster solution for this scam was cluster number 4. Solicitations Disguised as Invoices (217) from USPIS was placed in cluster 9 and was predicted to belong to cluster 3. The second cluster that this scam was predicted to belong to was cluster 2.

The Wilks' Lambda results in Table 75 in the Appendix confirm that each of discriminant functions are significant, all have p-values less than 0.05 (p = 0.00) and large Chi-square values. From the Eigenvalues table in Table 74 in the Appendix, it can be seen that the first function accounts for 31.5% of the total variation within the clustering of scam cases (E = 7.465). The seven-function solution accounts for 93% (E = 2.819) of variation.

4.4.7 DFA: Comparison Summary

The purpose of performing discriminant function analyses on the most promising results from the hierarchical clustering procedures was to identify which hierarchical model best partitioned scam cases into homogeneous groups. This was achieved by using the cluster membership results from six suitable hierarchical procedures as the dependent variables in six individual discriminant analysis procedures and the same static features used in the hierarchical procedure as the independent variables in the discriminant function analysis. The goal of the hierarchical procedure was to find the fewest clusters of scam cases containing the least number of scam cases per cluster. The goal of the discriminant function analysis was to test the accuracy of the cluster models, seeking at least a 95% level of accuracy. A secondary goal of the discriminant function analysis was to identify which static features most impact the prediction of scam cluster memberships, comparative results of the discriminant function analyses performed appear below in Table 23

Table 23: Summary Table of Cluster Model and its Level of Accuracy

Method	Clusters	Accuracy
Furthest neighbour, Jaccard	9	96.3
Furthest neighbour, Jaccard	8	95.7
Furthest neighbour, Jaccard	7	94.7
Furthest neighbour. Jaccard	6	92.9
Within groups, Jaccard	9	95.4
Within groups, Jaccard	8	93

The first discriminant function analysis performed was on the results from the nine cluster model of the furthest neighbour, Jaccard coefficient hierarchical clustering procedure. This model accurately predicted 96.3% of scam cluster memberships. The second discriminant procedure was performed on the results from the eight cluster model of the furthest neighbour, Jaccard coefficient hierarchical clustering procedure. This model accurately predicted 95.7% of scam cluster memberships. The third discriminant procedure performed was on the 7 cluster solution of the furthest neighbour, Jaccard coefficient hierarchical procedure. This model accurately predicted 94.7% of scam cluster memberships. The fourth discriminant procedure was performed on the results of the six cluster solution furthest neighbour, Jaccard coefficient hierarchical clustering analysis. This model accurately predicted 92.9% of scam cluster memberships. The fifth discriminant function analysis performed was on the results from the nine cluster model of the within groups linkage, Jaccard coefficient hierarchical clustering procedure. This method accurately predicted 95.4% of scam cluster memberships. The sixth and final discriminant function analysis was performed on the 8 cluster solution from the within groups linkage, Jaccard coefficient hierarchical clustering model. This method accurately predicted 93% of scam cluster memberships.

There are two models of interest which emerge from the hierarchical cluster results and are confirmed by the discriminant function analysis and these are the 8-cluster furthest neighbour Jaccard coefficient and the 7-cluster furthest neighbour Jaccard coefficient models. The 8-cluster model from the furthest neighbour and Jaccard method is selected as preferable a model because it contain one of the fewest number of clusters and less than 5% of variation in cluster placement results (p = 95.7%). The 7-cluster model from the same hierarchical method contains the fewest clusters while the amount of variation in scam group placement is only slightly greater than 5% (94.7%). This solution is of interest because if the error margin can be reduced to no more than 5% then this model would offer the best solution since it would become the solution with the fewest number of clusters with the least amount of acceptable variation. The seven cluster model was therefore re-tested with all non-significant static features removed.

With the non-significant static features removed, the model 7-cluster contains 68 predictor variables, all of which test significant to the model as can be seen in the Tests of Equality of Group Means that appear in Table 77 with Insignificant Features Removed in the Appendix.

There are thirteen scam entries that were predicted to belong to a different cluster than that assigned to them during the hierarchical clustering analysis; these results appear in Table 80 with Insignificant Features Removed in the Appendix. All of these scams however, were reassigned to the original cluster that they were placed within during the hierarchical procedure during the second stage of cluster membership prediction. These were scam number 19, Free Offers on the Internet from Scamwatch. This was predicted to belong to cluster 1 while it was placed into cluster 5 by the hierarchical procedure. Scam number 45, Credit Card Fraud from the IC3, was predicted to belong to cluster 1 while it was placed into cluster 5 by the hierarchical procedure. Scam number 46, Debt Elimination by the IC3, was predicted to belong to cluster 2 while it was placed into cluster 5 by the hierarchical procedure. Scam number 50, Identity Theft from the IC3, was predicted to belong to cluster 3 while it was placed into cluster 5 by the hierarchical procedure. Scam number 63, Phishing and Related Scams from the ABS, was predicted to belong to cluster 7 while it was placed into cluster 1 by the hierarchical procedure. Scam number 72, Get Rich Quick Scams from the OFT, was predicted to belong to cluster 3 while it was placed into cluster 7 by the hierarchical procedure. Scam number 82, Advance Fee Vacation Fraud from the ERG, was predicted to belong to cluster 2 while it was placed into cluster 1 by the hierarchical procedure. Scam number 202, Multilevel Marketing from the USPIS, was predicted to belong to cluster 3 while it was placed into cluster 7 by the hierarchical procedure. Scam number 207, Advance Fee Loan Scam from the USPIS, was predicted to belong to cluster 1 while it was placed into cluster 2 by the hierarchical procedure. Scam number 212, Unclaimed Income Tax Refund from the USPIS, was predicted to belong to cluster 2 while it was placed into cluster 1 by the hierarchical procedure. Scam number 220, Illegal Sweepstake Information from the USPIS, was predicted to belong to cluster 2 while it was placed into cluster 1 by the hierarchical procedure. Scam number 227, Home Improvement and Repair Fraud from the USPIS, was predicted to belong to cluster 1 while it was placed into cluster 2 by the hierarchical procedure. Lastly, scam number 261, Cold Calling from FIDO, was predicted to belong to cluster 7 while it was placed into cluster 1 by the hierarchical procedure.

Since all of the scams predicted to belong to a different group to that assigned to them during the hierarchical clustering analysis were predicted to belong to their originally derived group during the second phase of discriminant predictions, it is concluded that the seven-cluster model is suitable for homogeneous scam case clustering. The Wilks' Lambda results in Table 79 with Insignificant Features Removed and Table 78 with Insignificant Features Removed in the Appendix, further confirms this conclusion with 95% of the variation within the model being accounted for by the first six functions with all functions significant to the model (p = 0).

4.5 Summary

The purpose of the analyses performed here were to identify homogeneous subsets of scam cases. This was achieved through the use of hierarchical clustering analysis and discriminant function analysis. The aim of this research was to formulate clusters of scam cases derived from similarity matching principles based upon the purposely derived static features of scam descriptions. It was hypothesized that a smaller number of scam clusters could be found than the publicly acknowledged 38 which were recorded during the data collection phase. Hierarchical clustering was selected as the optimal choice for clustering analysis because it was unknown how many scam clusters there would be and the data was limited by its binary form. This method of clustering was also selected because of its natural tendency to find homogeneous subsets within a data set. Therefore, another aim of this research was to find the most reliable partitioning of scam cases which would allow for the homogeneous clustering of scam cases across the fewest number of clusters.

There were eight hierarchical clustering procedures performed on the purposefully derived scam static features. These included two binary distance methods for comparison and four linkage methods suitable for binary data, also for comparison. The binary distance measures compared were the Jaccard coefficient and the Simple Matching coefficient. The four linkage methods compared were the furthest neighbour, within groups linkage, between groups linkage, and nearest neighbour methods. The memberships of scam clusters were recorded for each analysis performed and dendrograms and bar charts tabulating the frequencies within each cluster were created. Conclusions were drawn from inspecting the dendrograms and bar charts from each round of analysis.

Conclusions were drawn from the cluster frequency results for each hierarchical clustering procedure. The key points reported upon were the clusters with the minimum and maximum frequencies and the range and shape of each distribution. A primary contributing factor to the conclusions drawn regarding the acceptability of a model or rejection of a model were due to the range of frequency solutions and the shape of the distribution. If a distribution contained a single cluster with 100 scam cases or greater, it was rejected and it was concluded that the model was unsuited to the homogeneous partitioning of scam cases. This was because a cluster with 100+ scam cases did not provide enough segmentation or discrepancy in scam cluster identification which would cause difficulty in further analysis. Since one of the goals of this research is to identify relatively homogeneous groups of scam cases, it is assumed that the ideal solution will be representative of the sample and scam cases will be relatively evenly distributed amongst the final number of identified scam clusters, thus removing all concern over limited or small cluster sample sizes. The results found from this investigation are general to the sample tested and this sample alone. Further scam gathering from a wider variety of sources is necessary to compile a data set large enough and comprehensive enough to use the results to make generalization of the wider population. This investigation is a pre-curser for such a study and as such is an investigation into the idea that divisively derived scam static features can be used as a the basis of scam structure for the identification of like and dislike cases of scam type using hierarchical clustering analysis. All results are pertinent to this sample only and should be interpreted with caution in the wider context.

The inspection of hierarchical clustering results concluded that the Jaccard distance coefficient was most suitable for the homogeneous partitioning of scam cases. The Simple Matching distance coefficient did not provide any suitable results for either linkage method tested. The nearest neighbour and within groups linkage methods were the most appropriate linkage method for partitioning scam cases into homogeneous subsets. The between groups and nearest neighbour methods did not proved any suitable results across either distance measure. The hierarchical clustering model that provided the most promising results was the furthest neighbour – Jaccard coefficient model. This model provided four cluster groupings that were selected for further analysis. These were the nine, eight, seven and six cluster solutions. The within groups – Jaccard coefficient model provided two cluster groupings that were selected for further analysis and these were the nine and eight cluster solutions.

Those hierarchical cluster models selected were further analysed by a discriminant function analysis. The goal of the discriminant function analysis was to assess the reliability of each model. A reliable model was defined as a model in which the *(n-1)* discriminant functions combined accounted for at least 95% of variability within the data. A model which accounted for at least 95% of variability could then be claimed to be at least 95% accurate. Another goal of the discriminant function analysis was to identify which scam static features were significant to cluster membership prediction.

The cluster solutions found to be accurate at least 95% of the time and with the fewest number of clusters were the 8-cluster and 7-cluster furthest neighbour, Jaccard coefficient models. With all insignificant static features removed, the 7-cluster model was found to be accurate 95% of the time. The number of static features significant to the accurate partitioning of scam cases into homogeneous subsets was 68. The following chapter, Chapter 7 discusses these results in further detail and delivers concluding remarks for each problem statement. The results from each of the 7-cluster and 8-cluster model outlined above are discussed and the resultant scam genres identified and labeled. The validation of the furthest neighbour, Jaccard coefficient hierarchical clustering model for scam-based research carries implications for the usability of text-based publicly accessible data in mixed methodological and quantitative analysis. These results also confirm the variability of scam descriptions across jurisdictions and provide evidence for the necessity of the standardisation of terminology.

Chapter 5: Conclusion

5.1 Discussion

The purpose of this research was to homogeneously group scam perpetrations into hierarchical groups of scam genres using clustering and discriminant function analysis of derived static features. Prenzler and Hays (2002) demonstrate that a hierarchy of fraud exists and it was inferred that a hierarchy if scams also exists. Similar to the methods of Holt and Graves (2007), data was derived from publicly available scam descriptions and through a bottom-up, grounded theoretical approach involving the manual content analysis of scam descriptive texts, scam static features were derived. Like Airoldi and Malin (2004) work on fraudulent intent detection in emails, hierarchical cluster analysis was used to identify homogeneous groups of scams by partitioning scams into similar clusters. The suitable cluster solutions were then tested for accuracy using discriminant function analysis. The residual effect of clustering scams by their descriptions is the standardisation of scam descriptions and identification of significant static features which can be used to confidently identify the type of scam a scam is.

5.1.1 Resolving the Research Problems

The first research objective was to cluster scam descriptions by partitioning them into scam genres. The research problem associated with this objective is the Homogeneous Grouping Problem which aimed to determine the suitability of hierarchical cluster analysis for the homogeneous partitioning of scam cases. The second research objective was to measure the effectiveness of using scam static features for the partitioning of scam cases into scam genres. The research problem associated with the second objective is the Static Feature Selection Problem which aimed to resolve the necessity of scam static features in determining scam case partitioning and scam genre membership as well as identify which static features most impact on scam case placement. The third research objective was to find the clustering model which best reduces scam cases into the fewest number of clusters with the least number of scam cases within each cluster which is accurate at least 95% of the time. The research problem associated with the third objective is the Minimum Cluster and Least Membership Problem.

5.1.2 Homogeneous Grouping Problem

To address the Homogeneous Grouping Problem, four linkage methods and two distance measures were tested and the results compared. The hierarchical model found to best partition scam cases into homogeneous groups using the static features as independent variables was the furthest neighbour, Jaccard coefficient model. Second to this was the within groups linkage, Jaccard coefficient model. Each model relying on the Simple Matching coefficient was not useful in partitioning scams as was the nearest neighbour and between groups linkage, Jaccard coefficient models. The furthest neighbour, Jaccard coefficient model was able to partition scam cases into homogeneous subsets while finding the least number of scam clusters with the fewest number of scam cases in each cluster.

The results found here suggest that the proposed combination of method and measure is most suited to explicit binary category membership data of the sample tested here.

5.1.3 Static Feature Selection Problem

Scam static features were derived from scam descriptive texts and these were used as independent variables for the homogeneous clustering of scam cases. The Static Feature Selection Problem explores the usefulness of using static features in determining scam membership and identifies those static features explaining the most variation in scam case assignment. This research problem also provides recommendations for the required number of scam static features in determining scam genre membership.

The first question of the Static Feature Selection Problem asked if static features could be used to determine the scam membership of scam descriptions. It is concluded that scam static features are an informative and useful data source for determining scam membership of scam perpetrations.

The most effective clustering model for achieving this, using scam static features as the clustering variable, is the furthest neighbour Jaccard coefficient hierarchical clustering model.

The second question (a) of the Static Feature Selection Problem asked how many static features were required to determine scam membership. Two final models were selected for comparison and these were the 8 and 7-cluster furthest neighbour Jaccard coefficient models. The recommendation for the most appropriate model is provided in the next section and without divulging too much detail it can be concluded for the Static Feature Selection Problem that 68 of the initial 82 scam static features are required for the successful assignment of scam cases to scam genres. The 68 scam static features are all significant to the model at the alpha = 5% level. Some features were more significant than others however, and this can be determined from the feature p-values and the accompanying F-value. A small p-value (less than 0.05) implies that the static feature is significant to the model and a large F-value infers greater contribution to cluster variation than a smaller F-value.

The third question (b) of the Static Feature Selection Problem required the identification of those static features most useful in determining scam case scam membership. There were two static features which were most important in identifying which cluster a scam case belonged to and these were what the scam offered (see the below table), - *employment* ($F = 243.588$, $p = 0.00$), and the role of the victim – *customer* ($F = 123.176$, $p = 0.00$). Following these two most significant static features were the goal of the scammer – *financial gain* ($F = 92.958$, $p =0.00$), the role of the victim – *un-associated* (F – 64.366, $p = 0.00$), goal of the scammer – *information* ($F = 45.412$, $p = 0.00$), and the method of scam introduction – *received* ($F =35.474$, $p = 0.00$). Below in Table 24 is a summary of the remaining significant static features.

Table 24: Summary Table of Frequencies for F-Range and P-Values

F-range	p-value	Frequency
20 - 29.99	0	9
10 - 19.99	0	17
0 - 9.99	<0.05	36

The most significant static features for identifying the type of scam a scam belongs to are what the scam offered, the role of the victim, the goal of the scammer, and the method of scam introduction. What can be concluded here is that these static features are crucial to the successful scam campaign. If a scammer were to be planning a new scam campaign, these are the priority features that would need to be known and addressed by the scammer. Before launching a campaign, the scammer must know what he/she seeks, they must have an end goal decided, knowing this, the scammer develops a scam campaign which will deliver the desired outcome. From knowledge of the desired outcome, the scammer must then decide on how he/she will introduce the scam to the target, further to this, contingencies would be made on how to reach as many targets as necessary to meet the intended goal.

The scammer must know how to get the target involved in the scam, to do this, the scam must offer something to the target and finally, before the campaign can be finalised, the scammer must know what role he/she plays in the campaign, and therefore, what role the target will play. By focusing on these primary static features alone, rephrasing them into questions and finding answers to these questions, useful and detailed information can be gained about the type of scam a scam might be and this would assist investigators and researchers alike in their quest for answers.

5.1.4 Minimum Cluster and Least Membership Problem

The third research objective aimed to find the clustering model which best condensed scam cases into the fewest number of clusters with the least number of scam cases within each cluster which would be accurate at least 95% of the time. The Minimum Cluster and Least Membership Problem is focused on determining which of the selected hierarchical models satisfy the minimum cluster, least membership and accuracy of 95% and this was achieved through applying discriminant function analysis on the results from the selected hierarchical models. Six hierarchical solutions were tested, four from the furthest neighbour, Jaccard coefficient analyses and 2 from the within groups linkage, Jacccard coefficient analyses.

The solutions for the number of scam clusters tested ranged from 6 to 9 and the results found suggest that either hierarchical model, purposefully clustering scam cases by static features into 9 clusters will achieve 96% accuracy in scam prediction. This is a good result because it can be immediately concluded that the number of scam genres that exist is 9 rather than the recorded 38 that were found during the data collection phase. However, the purpose of this research was to find as few clusters with as few scams in each cluster as possible with a 95% level of accuracy.

The 8-cluster furthest neighbour Jaccard coefficient hierarchical cluster model was accurate in scam case assignment 95.7% of the time. The first scam cluster detailed in Table 57 contains 21 scam cases. The second scam cluster detailed in Table 58 contains 74 scam cases and the third scam cluster detailed in Table 59 contains 51 scam cases. The fourth scam cluster detailed in Table 60 contains 22 scam cases and the fifth scam cluster in Table 61 contains 17 scam cases. The sixth scam cluster detailed in Table 62 contains 38 scam cases while the seventh scam cluster detailed in Table 63 contains 24 scam cases and the final scam cluster in Table 64 contains 30 scam cases.

Table 25: Scam Genre 1: 8 Cluster Model

Scam Name	Source	Country
Door to door	SW	Aus
Cheque overpayment	SW	Aus
Counterfeit cashiers check	IC3	USA
Fake clairvoyant	OFT	UK
Overpayment for sale of merchandise fraud	ERG	Can
Clairvoyant and psychic mailing scam	OFT	UK
Rolling lab schemes	FBI	USA
Check overpayment	OGO	USA
Multiple bidding	L2G2BT	USA
Counterfeit cashiers check	L2G2BT	USA
Overpayment scam	ACCC	Aus
Miracle cure	ACCC	Aus
Weight loss	ACCC	Aus
Door to door scam	ACCC	Aus
Business opportunities	ACCC	Aus
Solicitations disguised as invoices	USPIS	USA
Missing persons fraud	USPIS	USA
Cheque overpayment	SS	Aus
Fraudulent cheques or credit card scam	QPOL	Aus
Investment fraud	USPIS	USA
Internet extortion	IC3	USA

[2]

The first scam genre in Table 25 above contains a mixture of scam types from scams that are delivered in a face to face context such as door to door scams to those delivered over the Internet like multiple bidding and Internet extortion scams. The scams clustered into scam genre one come from a variety of sources and jurisdictions, Australian and the United States featuring equally. The feature which stands out the most as a commonality among these scams is the once off or up front payment to the scammer. It is suggested that the goal of the scammer for this group of scams could be financial gain and this scam genre can be labelled **Financial Gain Scams through Limited Transaction Periods**.

The second scam genre in Table 26 below contains a greater mixture of scam types than that seen in scam genre one. There are a number of advance fee scams, Nigerian 419 scams and lottery scams. In this scam genre appear charity scams and 900 telephone number scams as well. Scams sourced from the United States appear more frequently than any other country in scam genre two, followed by Australia then Canada and the UK equally and with this wide range of scam types it is difficult to identify one or two distinct similarity properties which could be used to label this scam genre.

---

[2] Scamwatch (SW, Internet Crime Complaint Center (IC3), Office of Fair Trading (OFT), Environics Research Group (ERG), Federal Bureau of Investigation (FBI), On guard online (OGO), Looks too good to be true (L2G2BT), Australian Competition and Consumer Commission (ACCC), United States Postal Inspectors Service (USPIS), Scam smart (SS), and Queensland Police Service (QPOL)

Scam genre three in Table 27 below also contains a range of scam types sourced equally from Australia and the US and with a minor contribution from Canada and the UK. For those scams placed into scam genre three, it is difficult to infer a title because there is such wide variety in scam type present.

Table 26: Scam Genre 2: 8-Cluster Model

Scam Name	Source	Country	Scam Name	Source	Country
Charity	SW	Aus	Advance fee scam	L2G2BT	USA
Dating and romance	SW	Aus	Charities fraud	L2G2BT	USA
Fax back	SW	Aus	Nigerian 419	L2G2BT	USA
Spam offers	SW	Aus	Foreign lottery	L2G2BT	USA
Upfront payment	SW	Aus	Sweepstakes and prizes scams	L2G2BT	USA
Nigerian 419	SW	Aus	Lottery	ACCC	Aus
Lottery and sweepstakes	SW	Aus	Fake prize	ACCC	Aus
Unexpected prizes	SW	Aus	Chain letters	ACCC	Aus
Chain letters	SW	Aus	Nigerian scam	ACCC	Aus
Lotteries	IC3	USA	Inheritance scam	ACCC	Aus
Nigeria letter 419	IC3	USA	Dating and romance	ACCC	Aus
Advance fee fraud	ABS	Aus	Distributorship and franchise fraud	USPIS	USA
Chain letters	ABS	Aus	900 telephone number fraud	USPIS	USA
Lottery	ABS	Aus	Advance fee loan schemes	USPIS	USA
Advance fee	OFT	USA	Charity fraud	USPIS	USA
International sweepstakes	OFT	USA	Chain letters	USPIS	USA
Prize draw pitch	OFT	USA	Free prize scheme	USPIS	USA
Bogus lottery	OFT	USA	Foreign lotteries	USPIS	USA
High pressure sales pitch vacation	ERG	Can	Telemarketing fraud	USPIS	USA
Prize lottery and sweepstakes	ERG	Can	Home improvement and repair fraud	USPIS	USA
West African 419	ERG	Can	Phony in heritance scam	USPIS	USA
Advance fee loan fraud	ERG	Can	Prison pen pal money order scam	USPIS	USA
Upfront fee for credit card fraud	ERG	Can	Nigerian	SS	Aus
Prize draw and sweepstake	OFT	UK	Lottery prize	SS	Aus
Foreign lottery scams	OFT	UK	Holiday prize	SS	Aus
Premium rate telephone prize scams	OFT	UK	Internet bride	SS	Aus
African advance fee frauds foreign money making	OFT	UK	Inheritance scam	SS	Aus
Bogus holiday club scams	OFT	UK	Churches	SS	Aus
Telemarketing fraud	FBI	USA	Bowling clubs	SS	Aus
Nigerian or 419	FBI	USA	Hitman	SS	Aus
Advance fee scheme	FBI	USA	Dating dowry and romance	SS	Aus
Nigerian email scam	OGO	USA	Donation	SS	Aus
Foreign lotteries	OGO	USA	Nigerian letter and advance fee fraud	FIDO	Aus
Pay in advance credit offers	OGO	USA	Lottery scams	FIDO	Aus
Debt relief	OGO	USA	Request to use bank account	QPOL	Aus
Cross border fraud	L2G2BT	USA	Online relationship	QPOL	Aus
Romance scheme	L2G2BT	USA	Charity scam	QPOL	Aus

Table 27: Scam Genre 3: 8-Cluster Model

Scam Name	Source	Country
Psychic and clairvoyant	SW	Aus
Office supply	SW	Aus
Directories and advertising	SW	Aus
Fake online pharmacies	SW	Aus
Weight loss	SW	Aus
Miracle cures	SW	Aus
Domain name renewal	SW	Aus
Cold calling	SW	Aus
Financial advice	ABS	Aus
Pyramid schemes	ABS	Aus
Credit or bank card	ABS	Aus
Bogus investment	OFT	UK
Miracle health cure	OFT	UK
Bogus health product cure	ERG	Can
Investment fraud	ERG	Can
Advance fee vacation fraud	ERG	Can
Miracle health and slimming cure scams	OFT	UK
High risk investment scams	OFT	UK
Letter of credit fraud	FBI	USA
Prime bank note	FBI	USA
Weight loss claims	OGO	USA
Cure all products	OGO	USA
Pharmacy fraud	L2G2BT	USA
Investments fraud	L2G2BT	USA
Health and diet scams	USC	USA
Cold calling	ACCC	Aus
Share promotions and hot tips	ACCC	Aus
Gambling software	ACCC	Aus
Fake online pharmacies	ACCC	Aus
Psychic or clairvoyant	ACCC	Aus
Small business scams	ACCC	Aus
Directory entry unauthorised advertising	ACCC	Aus
Mystery shopper scam	USPIS	USA
Credit card fraud	USPIS	USA
Child support collection scheme	USPIS	USA
Social security schemes	USPIS	USA
Unclaimed income tax refund	USPIS	USA
Unclaimed funds scheme	USPIS	USA
Property tax exemption scheme	USPIS	USA
Cut rate health insurance fraud	USPIS	USA
Oil and gas investment fraud	USPIS	USA
Land fraud	USPIS	USA
Illegal sweepstakes information	USPIS	USA
Government look alike mail	USPIS	USA
Free vacation scams	USPIS	USA
Receipt for unsolicited merchandise	USPIS	USA
Fraudulent health and medical products	USPIS	USA
Astrology psychic and clairvoyant	SS	Aus
Share trading	SS	Aus
Cold calling	FIDO	Aus
Fake debt invoices	FIDO	Aus

Table 28: Scam Genre 4: 8-Cluster Model

Scam Name	Source	Country
Business opportunity	SW	Aus
Guaranteed employment and income	SW	Aus
Work from home	SW	Aus
Transferring money for someone else	SW	Aus
Employment or business opportunities	IC3	USA
Reshipping	IC3	USA
Third party receiver of funds	IC3	USA
Employment or work from home	ERG	Can
Cheque cashing money transfer job fraud	ERG	Can
Work at home and business opportunity scams	OFT	UK
Work at home scams	OGO	USA
Job scams	L2G2BT	USA
Counterfeit money orders	L2G2BT	USA
Bogus business opportunities	USC	USA
Work from home	ACCC	Aus
Guaranteed employment and income	ACCC	Aus
Phony job opportunities	USPIS	USA
Postal job scam	USPIS	USA
Work at home schemes	USPIS	USA
Employment work from home	SS	Aus
Money transfer	SS	Aus
Fake job email or money transfer schemes	FIDO	Aus

Scam genre four in Table 28 above is composed predominantly of employment-based scams from working at home to money transfer and job opportunity scams. While only one scam from the UK and two from Canada appear in scam genre four, Australia and the US are featured predominantly. This scam genre can be titled **Participation through Income Based Scenarios**.

Scam genre five in Table 29 is made up of phone-based and advance fee scams. This scam genre contains scam cases from Australia and the United Kingdom alone. The most common theme emerging for this scam genre is the **Financial Gain through Legitimate Appearing Scenarios**. A missed call and text message scam is successful because the victim is fooled into responding to the missed communication and is charged excessively for their call back. These fees and charges do not appear until the phone bill is received and may not even be recognised as an over-charge because not all phone users check their bill statements. Another example of scammer financial gain through a legitimate appearing scenario is the bogus model casting agency scam. In this scam the victim is a client of a fraudulent modelling or talent agency. The victim may have been approached by a talent scout for the agency or the victim may have submitted their portfolio for consideration to the phony agency. In either situation, a scenario emerges which appears to be a legitimate situation requiring the payment of fees.

Similarly to scam genre 5, for scam genre 6, found in Table 30, only scams sourced from Australia and the United States are featured. A common theme emerging is the type of scam; either semantic or syntactic, and the target of the scam which is information. For many of the scams in this scam genre, the scam is syntactically driven and for nearly all of the scams, the goal of the scam is to gather information. This scam genre can be titled **Information Gathering through Technology Based Tactics**.

Table 29: Scam Genre 5: 8-Cluster Model

Scam Name	Source	Country
SMS Competition and trivia	SW	Aus
Missed calls and text messages from unknown numbers	SW	Aus
Ring tone	SW	Aus
Modem jacking	SW	Aus
Superannuation	SW	Aus
Premium rate prize draw	OFT	UK
Property investment scams	OFT	UK
Internet dialer scams	OFT	UK
Bogus vanity publishers	OFT	UK
Bogus invention promotions	OFT	UK
Bogus model and casting agencies	OFT	UK
Loan scams	OFT	UK
Missed call	ACCC	Aus
Text message	ACCC	Aus
SMS Competition and trivia	ACCC	Aus
Faxback	ACCC	Aus
Office supply	ACCC	Aus

Table 30: Scam Genre 6: 8-Cluster Model

Scam Name	Source	Country
Spyware and key loggers	SW	Aus
Free offers on the internet	SW	Aus
Credit card	SW	Aus
Phony fraud alerts	SW	Aus
Requests for your account information	SW	Aus
Credit card fraud	IC3	USA
Debt elimination	IC3	USA
Identity theft	IC3	USA
Phishing or spoofing	IC3	USA
Spam	IC3	USA
Phishing and related	ABS	Aus
Identity theft	ABS	Aus
Impersonation or identity fraud	FBI	USA
Phishing	OGO	USA
Hacking	L2G2BT	USA
Identity theft	L2G2BT	USA
Phishing or spoofing	L2G2BT	USA
Spam	L2G2BT	USA
Spyware	L2G2BT	USA
Discount software offers	USC	USA
Phishing emails	USC	USA
Trojan horse email	USC	USA
Virus generated email	USC	USA
Phishing	ACCC	Aus
Fake fraud alerts	ACCC	Aus
Spam	ACCC	Aus
Malicious software	ACCC	Aus
Identity theft	SS	Aus
Phishing	SS	Aus
Software	SS	Aus
Virus	SS	Aus
Trojan	SS	Aus
Ransom ware	SS	Aus
Spyware	SS	Aus
Malware	SS	Aus
Fake bank emails	FIDO	Aus
Social networking fraud	FIDO	Aus
Identity theft	FIDO	Aus

Table 31: Scam Genre 7: 8-Cluster Model

Scam Name	Source	Country
Online auction and shopping	SW	Aus
Card skimming	SW	Aus
Product misrepresentation	IC3	USA
Non delivery	IC3	USA
Auction fraud Romania	IC3	USA
Parcel courier email schemes	IC3	USA
Escrow services fraud	IC3	USA
Bill for unsuitable merchandise	ERG	Can
Medical equipment fraud	FBI	USA
Services not performed	FBI	USA
Medicare fraud	FBI	USA
Debt elimination	L2G2BT	USA
Non delivery	L2G2BT	USA
Misrepresentation	L2G2BT	USA
Triangulation	L2G2BT	USA
Fee stacking	L2G2BT	USA
Black market or counterfeit goods	L2G2BT	USA
Shill bidding	L2G2BT	USA
International auction fraud	L2G2BT	USA
Escrow services scam	L2G2BT	USA
Card skimming	ACCC	Aus
Online auction and shopping	ACCC	Aus
Ringtone	ACCC	Aus
Online classifieds	SS	Aus

Table 32: Scam Genre 8: 8-Cluster Model

Scam Name	Source	Country
Identity theft	SW	Aus
Computer prediction software	SW	Aus
Investment seminars and real estate	SW	Aus
Share promotions and hot tips	SW	Aus
Pyramid schemes	SW	Aus
Investment fraud	IC3	USA
Ponzi or pyramid	IC3	USA
Get rich quick	OFT	UK
Bogus racing tipster	OFT	UK
Pyramid selling and chain letter scams	OFT	UK
Internet matrix scheme scams	OFT	UK
redemption or straw men or bond	FBI	USA
Ponzi scheme	FBI	USA
Pyramid schemes	FBI	USA
Investments schemes	OGO	USA
Ponzi or pyramid	OGO	USA
419 advance fee fraud	L2G2BT	USA
Pyramid schemes	ACCC	Aus
Investment seminar	ACCC	Aus
Charity	ACCC	Aus
Multilevel marketing	USPIS	USA
Affinity fraud	SS	Aus
Pyramid	SS	Aus
Ponzi	SS	Aus
Courses and seminars	SS	Aus
Pump and dump	FIDO	Aus
Pyramid schemes	FIDO	Aus
Ponzi scheme	FIDO	Aus
Affinity fraud	FIDO	Aus
Business opportunity	QPOL	Aus

Scam genre seven, in Table 31, above contains scams that are retail-based, predominantly in online auction situations from online auction and shipping to non delivery and misrepresentation. Most of the scams in this scam genre were sourced from the United States. Card skimming and ringtone scams appear in this scam genre also and this could be attributed to the customer-seller based relationship described in the original scam descriptions. This scam genre can be titled **Financial Gain through Retail Transactions**.

The final scam genre for the furthest neighbour Jaccard coefficient 8-cluster model is scam genre 8 found in Table 32, below. While there appears to be a mixture of scam types within this genre, a commonality emerges and this is the concept of investment. This scam genre contains pyramid, Ponzi, get rich quick, and betting scams, including seminar and business opportunity scams., and most of the scams found within this scam genre were sourced from Auastralia. Identity theft is also prominent in this scam genre and for these reasons; this scam genre can be labelled **Financial Gain and Information Gathering through Investment Opportunity**.

A seven-cluster solution provided by the furthest neighbour, Jaccard coefficient model achieved 94.7% accuracy, not quite reaching the required 95% level. By removing the non-significant static features, those which did not contribute significantly to the placement of scam cases into clusters or significantly account for variability among clusters were removed and the discriminant function procedure re-run. It was concluded that discriminant function analysis is useful in determining reliability of hierarchical models, it was also concluded that 68 scam static features were necessary in determining scam memberships, not the entire sample of 82. Finally, it was concluded that the fewest number of clusters with the least number of scam memberships, inferring homogeneity across clusters and among cases was 7 and scam cases could be accurately allocated to a scam genre (cluster) 95% of the time using the furthest neighbour, Jaccard coefficient hierarchical clustering model.

The first scam genre contains 72 scam cases, these are listed in Table 33. This scam genre is made up of scam cases that involve the most basic forms of trickery. These involve scams that are not necessarily thorough in planning and detail. Victims falling for scam genre one scams would take people and communications at face value and not expend time or energy on investigating scam claims or the people behind them. These scams target the individual or company for once off transactions initially and where possible, if there were potential for the scam to be extended to elicit more funds from the victim, this would be pursued. Scam genre 1 contains scams that are at the most basic level after the victim's money. Door to door scams often involve the soliciting of services that are paid for and never performed. Psychic and clairvoyant scams also involve the soliciting of services or merchandise that is paid for and is not what it had promised to be. Cheque overpayment scams involve the overpayment for a purchase and a request for the balance to be wired back. In this situation, the cheque is fraudulent and the scammer walks away with the victim's money. Financial advice scams involve soliciting financial advice for a fee. Whether or not the advice is useful is irrelevant since the victim has just paid a scammer and the scammer has walked away with their money and possibly their personal and private details to use in a future scam. Similarity among scams found in scam genre one emerge, the most significant is the payment of funds to the scammer, for this reason, scam genre one has been titled **Financial Gain through Low Level Trickery**.

Table 33: Scam Genre 1 – Financial Gain through Low Level Trickery

Scam Name	Source	Country	Scam Name	Source	Country
Door to door	SW	Aus	Cold calling	ACCC	Aus
Psychic & clairvoyant	SW	Aus	Share promotions & hot tips	ACCC	Aus
Office supply	SW	Aus	Gambling software	ACCC	Aus
Directories & advertising	SW	Aus	Overpayment	ACCC	Aus
Fake online pharmacies	SW	Aus	Miracle cures	ACCC	Aus
Weight loss	SW	Aus	Weight loss	ACCC	Aus
Miracle cures	SW	Aus	Fake online pharmacies	ACCC	Aus
Domain name renewal	SW	Aus	Psychic & clairvoyant	ACCC	Aus
Cheque overpayment	SW	Aus	Door to door	ACCC	Aus
Cold calling	SW	Aus	Business opportunities	ACCC	Aus
Counterfeit cashiers check	IC3	USA	Small business	ACCC	Aus
Internet extortion	IC3	USA	Direct entry unauthorised advertising	ACCC	Aus
Financial advice	ABS	Aus	Mystery shopper	USPIS	USA
Pyramid schemes	ABS	Aus	Credit card fraud	USPIS	USA
Credit & bank card	ABS	Aus	Child support collection scheme	USPIS	USA
Fake clairvoyant	OFT	UK	Social security schemes	USPIS	USA
Bogus investment	OFT	UK	Unclaimed income tax refund	USPIS	USA
Miracle health cure	OFT	UK	Unclaimed funds	USPIS	USA
Bogus health product	ERG	Can	Property tax exemption	USPIS	USA
Investment fraud	ERG	Can	Cut rate health insurance	USPIS	USA
Advance fee vacation fraud	ERG	Can	Investment fraud	USPIS	USA
Overpayment for sale of merchandise	ERG	Can	Solicitations disguised as invoices	USPIS	USA
Miracle health & slimming	OFT	UK	Oil & gas investment	USPIS	USA
Clairvoyant & psychic mailing	OFT	UK	Land fraud	USPIS	USA
High risk investment	OFT	UK	Illegal sweepstakes	USPIS	USA
Rolling labs	FBI	USA	Government look alike mail	USPIS	USA
Letter of credit fraud	FBI	USA	Free vacation scams	USPIS	USA
Prime bank note	FBI	USA	Receipt for unsolicited merchandise	USPIS	USA
Weight loss claims	OGO	USA	Missing persons	USPIS	USA
Cure all products	OGO	USA	Fraudulent health & medical products	USPIS	USA
Check overpayment	OGO	USA	Astrology psychic & clairvoyant	SS	Aus
Pharmacy fraud	L2G2BT	USA	Cheque overpayment	SS	Aus
Investments fraud	L2G2BT	USA	Share trading	SS	Aus
Multiple bidding	L2G2BT	USA	Cold calling	FIDO	Aus
Counterfeit cashiers check	L2G2BT	USA	Fake debt invoices	FIDO	Aus
Health & diet scams	USC	USA	Fraudulent cheques & credit cards	QPOL	Aus

The second scam genre contains 74 scam cases and these are listed below in Table 34. This scam genre is made up of scam cases that involve complex planning and detail. These scams hinge on the opportunistic nature of the general public as well as the scammer. In this sense a common bond is formed between the scammer and their victims and that is opportunity. The first scam in scam genre 2 is the charity scam. This scam relies on the poverty and necessity of others, it also comes about when natural, or man made disaster strikes. These scams rely on assumed public knowledge of a cohort of individuals or a global tragedy. They are story based scams and offer to their victims the opportunity to make a difference in the world through financial assistance. The ultimate goal of the scams found in scam genre 2 is money, the same as scam genre 1 however, the method of realising this goal is different. It would be worth investigating the dollar amounts lost to those scams found in scam genre 2 and compare them to scam genre 1 because it is suspected that scam genre 2 scams elicit greater amounts in funds while scam genre 1 elicits greater quantities of victims. It is interesting to see unexpected prizes and chain letters grouped together with charity scams and Nigerian 419 scams. This suggests some similarity in scam perpetrations; further investigation might prove useful in determining on what grounds these scams are alike. It may be due to the story – based nature of all of these scams. Another goal which manifests in dating and romance scams, Nigerian 419 scams, and even spam offers is the collection of personal or private information. For these reasons, scam genre two is titled **Financial Gain and Information Gathering through Developed Story Based Applications**.

Table 34: Scam Genre 2 – Financial Gain and Information Gathering Through Developed Story Based Applications

Scam Name	Source	Country	Scam Name	Source	Country
Charity	SW	Aus	Advance fee scam	L2G2BT	USA
Dating & romance	SW	Aus	Charities fraud	L2G2BT	USA
Fax back	SW	Aus	Nigerian 419	L2G2BT	USA
Spam offers	SW	Aus	Foreign lottery	L2G2BT	USA
Upfront payment	SW	Aus	Sweepstakes & prizes	L2G2BT	USA
Nigerian 419	SW	Aus	Lottery	ACCC	Aus
Lottery & sweepstakes	SW	Aus	Fake prize	ACCC	Aus
Unexpected prizes	SW	Aus	Chain letters	ACCC	Aus
Chain letters	SW	Aus	Nigerian scam	ACCC	Aus
Lotteries	IC3	USA	Inheritance scam	ACCC	Aus
Nigerian letter 419	IC3	USA	Dating & romance	ACCC	Aus
Advance fee fraud	ABS	Aus	Distributorship & franchise fraud	USPIS	USA
Chain letters	ABS	Aus	900 telephone numbers	USPIS	USA
Lottery	ABS	Aus	Advance fee loan schemes	USPIS	USA
Advance fee	OFT	UK	Charity fraud	USPIS	USA
International sweepstakes	OFT	UK	Chain letters	USPIS	USA
Prize draw pitch	OFT	UK	Free prize schemes	USPIS	USA
Bogus lottery	OFT	UK	Foreign lotteries	USPIS	USA
High pressure sales pitch vacation	ERG	Can	Telemarketing fraud	USPIS	USA
Prize lottery & sweepstakes	ERG	Can	Home improvement & repair	USPIS	USA
West African 419	ERG	Can	Phony inheritance	USPIS	USA
Advance fee loan	ERG	Can	Prison pen pal money order scam	USPIS	USA
Upfront fee for credit card	ERG	Can	Nigerian	SS	Aus
Prize draw & sweepstakes	OFT	UK	Lottery prizes	SS	Aus
Foreign lottery	OFT	UK	Holiday prizes	SS	Aus
Premium rate telephone prize	OFT	UK	Internet bride	SS	Aus
African advance fee frauds foreign mon	OFT	UK	Inheritance scam	SS	Aus
Bogus holiday club	OFT	UK	Churches	SS	Aus
Telemarketing	FBI	USA	Bowling clubs	SS	Aus
Nigerian or 419	FBI	USA	Hit man	SS	Aus
Advance fee scheme	FBI	USA	Dating dowry & romance	SS	Aus
Nigerian email	OGO	USA	Donation	SS	Aus
Foreign lotteries	OGO	USA	Nigerian letter & advance fee fraud	FIDO	Aus
Pay in advance credit offers	OGO	USA	Lottery scams	FIDO	Aus
Debt relief	OGO	USA	Request to use bank account	QPOL	Aus
Cross border fraud	L2G2BT	USA	Online relationship	QPOL	Aus
Romance scheme	L2G2BT	USA	Charity scam	QPOL	Aus

The third scam genre contains 22 scam cases and these appear in Table 35. This scam genre is made up of scam cases that involve complex planning and detail, similar to that found in scam genre 2. This scam genre however, targets the individual in the sense that it seeks participation from its victims. Each scam listed in scam genre three involves a level of victim 'employment' in which the victim participates in a scam which is normally a laundering scam and for their participation they are financially rewarded. These scams can often lead to identity theft since in becoming involved in one of these scams; the victim may have been an applicant for what they had believed was an authentic employment opportunity. With their application, the victim would have supplied the scammer/s with a full working and educational history, full name and date of birth as well as bank account details. For these reasons, scam genre three has been titled **Participation and Information Gathering through Employment Based Strategies**.

Table 35: Scam Genre 3 – Participation and Information Gathering through Employment Based Strategies

Scam Name	Source	Country
Business opportunity	SW	Aus
Guaranteed employment & income	SW	Aus
Work from home	SW	Aus
Transferring money for someone else	SW	Aus
Employment or business opportunities	IC3	USA
Re-shipping	IC3	USA
Third party receiver of funds	IC3	USA
Employment work from home	ERG	Can
Cheque cashing money transfer job fraud	ERG	Can
Work at home & business opportunity scams	OFT	UK
Work at home scams	OGO	USA
Job scams	L2G2BT	USA
Counterfeit money orders	L2G2BT	USA
Bogus business opportunities	USC	USA
Work from home	ACCC	Aus
Guaranteed employment	ACCC	Aus
Phony job opportunities	USPIS	USA
Postal job scams	USPIS	USA
Work at home schemes	USPIS	USA
Employment work from home	SS	Aus
Money transfer	SS	Aus
Fake job email or money transfer schemes	FIDO	Aus

The fourth scam genre contains 17 scam cases and these appear below in Table 36. This scam genre is made up of scam cases that require victim call backs or responses to be successful. The scams found here are different to those seen in scam genre one, two, and three. Most of these scams rely on alternative technologies to that of the Internet and World Wide Web for dissemination. There are a mixture of scams here that aim to trick the victim into responding and thus facing un-realised charges. Regardless of the method of the scam, or the role of the victim, this scam genre contains scams that aim to make money from the victim in ways that would seem necessary or pertinent to the situation. For this reason, scam genre four is titled **Financial Gain through Implied Necessary Obligation**.

Table 36: Scam Genre 4 – Financial Gain through Implied Necessary Obligation

Scam Name	Source	Country
SMS competition & trivia	SW	Aus
Missed calls & text messages from unknown n	SW	Aus
Ring tone	SW	Aus
Modem jacking	SW	Aus
Superannuation	SW	Aus
Premium rate prize draw	OFT	UK
Property investment	OFT	UK
Internet dialer	OFT	UK
Bogus vanity publishers	OFT	UK
Bogus invention promotions	OFT	UK
Bogus model & casting agencies	OFT	UK
Loan scams	OFT	UK
Missed calls	ACCC	Aus
Text messages	ACCC	Aus
SMS competition & trivia	ACCC	Aus
Faxback	ACCC	Aus
Office supply	ACCC	Aus

The fifth scam genre contains 38 scam cases and these appear below in Table 37. This scam genre is made up of scam cases that involve high level knowledge of how systems operate. This scam genre contains those scams that are syntactically driven such as spyware and key logger scams. This scam genre also contains scams that seek information for the purpose of identity theft and credit/debit card fraud. The reason why syntactic scams using spyware and key loggers are clustered along with identity theft and credit/debit card scams is because syntactic attacks are dispersed with the goal of gathering victim identity credentials or other forms of information. Therefore spyware and key logging scams are a tool for the success of information gathering scams such as identity theft and credit/debit card scams. Also found in scam genre 5 are phishing scams. These scams are also synonymous with identity theft and credit/debit card fraud which was described in further detail in the literature review section. For these reasons, scam genre five is titled **Information Gathering through Apparently Authentic Appeals**.

Table 37: Scam Genre 5 – Information Gathering through Apparently Authentic Appeals

Scam Name	Source	Country
Spyware & key-loggers	SW	Aus
Free offers on the internet	SW	Aus
Credit card	SW	Aus
Phony fraud alerts	SW	Aus
Requests for account information	SW	Aus
Credit card fraud	IC3	USA
Debt elimination	IC3	USA
Identity theft	IC3	USA
Phishing & spoofing	IC3	USA
Spam	IC3	USA
Phishing & related	ABS	Aus
Identity theft	ABS	Aus
Impersonation or identity fraud	FBI	USA
Phishing	OGO	USA
Hacking	L2G2BT	USA
Identity theft	L2G2BT	USA
Phishing & spoofing	L2G2BT	USA
Spam	L2G2BT	USA
Spyware	L2G2BT	USA
Discount software offers	USC	USA
Phishing email	USC	USA
Trojan horse email	USC	USA
Virus generated email	USC	USA
Phishing	ACCC	Aus
Fake fraud alerts	ACCC	Aus
Spam	ACCC	Aus
Malicious software	ACCC	Aus
Identity theft	SS	Aus
Phishing	SS	Aus
Software	SS	Aus
Virus	SS	Aus
Trojan	SS	Aus
Ransom-ware	SS	Aus
Spyware	SS	Aus
Malware	SS	Aus
Fake bank emails	FIDO	Aus
Social networking fraud	FIDO	Aus
Identity theft	FIDO	Aus

The sixth scam genre contains 24 scam cases and these appear in Table 38. This scam genre is made up of scam cases that involve and incorporate the roles of seller and buyer in the scam description. These scams are all transaction based auction – retailer style scams. Internet auction scams were described in detail in the literature review section of this research where five auction scams were identified: shill bidding, bid shielding, merchandise non-delivery, payment non-delivery, and product authenticity. All of these pre-identified Internet auction scams appear in their many guises below. The goal of these scams is financial gain which is achieved through various versions and applications of similarly styled scams. These scams are well researched and developed even though the victim and scammer only communicate for a short period of time. For these reasons, scam genre six is titled **Financial Gain through Merchant and Customer Based Exploitation**.

Table 38: Scam Genre 6 – Financial Gain through Merchant and Customer Based Exploitation

Scam Name	Source	Country
Online auction & shopping	SW	Aus
Card skimming	SW	Aus
Product misrepresentation	IC3	USA
Non delivery	IC3	USA
Auction fraud Romania	IC3	USA
Parcel courier email scheme	IC3	USA
Escrow services fraud	IC3	USA
Bill for unsuitable merchandise	ERG	Can
Medical equipment fraud	FBI	USA
Services not performed	FBI	USA
Medicare fraud	FBI	USA
Debt elimination	L2G2BT	USA
Non-delivery	L2G2BT	USA
Misrepresentation	L2G2BT	USA
Triangulation	L2G2BT	USA
Fee stacking	L2G2BT	USA
Black market or counterfeit goods	L2G2BT	USA
Shill bidding	L2G2BT	USA
International auction fraud	L2G2BT	USA
Escrow services scam	L2G2BT	USA
Card skimming	ACCC	Aus
Online auctions & shopping	ACCC	Aus
Ringtone	ACCC	Aus
Online classifieds	SS	Aus

The seventh and final scam genre contains 30 scam cases and these appear in Table 39. This scam genre is made up of scam cases that involve the exploitation of investment opportunities. This scam genre contains a mixture of scam types including Ponzi and pyramid, identity theft, computer prediction software, investment seminars, charity fraud, affinity fraud, get rich quick scams and 419 advance fee fraud. Without further detailed analysis of the inter-connected nature of scam static features to pin point the reason behind this, the presence of this mixture of scam titles is interpreted as hinging on the suggestion of investment opportunities within each scam case. The scams within scam genre 6 are marketed as money making opportunities, whether through investment, business opportunity, shares or gambling. However, the goal of the scammer is financial gain and in some instances this extends to information gathering. For these reasons, scam genre seven is titled **Financial Gain and Information Collection through Marketing Opportunities**.

Table 39: Scam Genre 7 – Financial Gain and Information Collection through Marketing Opportunities

Scam Name	Source	Country
Identity theft	SW	Aus
Computer prediction software	SW	Aus
Investment seminars & real estate	SW	Aus
Share promotions & hot tips	SW	Aus
Pyramid schemes	SW	Aus
Investment fraud	IC3	USA
Ponzi or pyramid	IC3	USA
Get rich quick	OFT	UK
Bogus racing tipster	OFT	UK
Pyramid selling & chain letter	OFT	UK
Internet matrix scams	OFT	UK
Redemption strawmen or bond	FBI	USA
Ponzi scheme	FBI	USA
Pyramid schemes	FBI	USA
Investment schemes	OGO	USA
Ponzi or pyramid	L2G2BT	USA
419 advance fee fraud	USC	USA
Pyramid scheme	ACCC	Aus
Investment seminar	ACCC	Aus
Charity	ACCC	Aus
Multilevel marketing	USPIS	USA
Affinity fraud	SS	Aus
Pyramid	SS	Aus
Ponzi	SS	Aus
Courses & seminars	SS	Aus
Pump & dump	FIDO	Aus
Pyramid schemes	FIDO	Aus
Ponzi scheme	FIDO	Aus
Affinity fraud	FIDO	Aus
Business opportunity	QPOL	Aus

While the 8-cluster model was slightly more accurate than the 7-cluster model, the 7-cluster model achieved better partitioning of scam cases into groups called scam genres that could be confidently titled based upon the types of scams receiving assignment to them. The 7-cluster model also achieved the minimum accuracy requirement of 95% accuracy with all non-significant static features removed. The second and third scam genres of the 8-cluster model could not be titled since it contained such a wide mixture of scam types while all scam genres of the 7-cluster model could be titled. The final model which satisfies the requirements of the Minimum Cluster and Least Membership Problem is the 7-cluster furthest neighbour Jaccard coefficient hierarchical clustering model. This model successfully partitions scam cases into the fewest scam genres with the least number of scam cases per scam genre with 95% accuracy and requires 68 of the 82static features to do so.

5.2 Summary

Two hundred and seventy seven individual scam cases and 82 purposely derived scam static features belonging to 38 separate source classified scam genre categories were analysed using an unsupervised agglomerative furthest neighbor, Jaccard coefficient hierarchical clustering model which was verified and tested for reliability by a discriminant function analysis. This method achieved 95% accuracy in partitioning scam cases into scam genres. The 38 source classified scam genres were reduced down to only 7 scam genres which were Financial Gain through Low Level Trickery, Financial Gain and Information Gathering through Developed Story Based Applications, Participation and Information Gathering through Employment Based Strategies, Financial Gain through Implied Necessary Obligation, Information Gathering through Apparently Authentic Appeals, Financial Gain through Merchant and Customer Auction Based Exploitation, and Financial Gain and Information Collection through Marketing Opportunities. It was discovered that only 68 of the 82 scam static features were required to achieve a 95% level of accuracy in scam membership and the most prominent of these static features were what the scam offered, the role of the victim, the goal of the scammer, and the method of scam introduction.

It is concluded that hierarchical clustering using the furthest neighbour and Jaccard coefficient is a reliable method of clustering scam static features and that scam static feature derived from publicly available scam descriptions are a useful source of information for scam investigation. It is also concluded that scams are currently over classified within current literature and that only 7 scam types or scam genres exist compared to the 38 recorded source-classified scam categories.

5.3 Future Work

Future work in this area would involve the collection of a larger sample of data including scam descriptions from non-English speaking origins. From this the reliability of the scam static features, hierarchical clustering model and the seven - cluster scam genre model revealed in this research could be further verified. Building from the methodology applied within this research, a case study analysis of scam perpetrations from initial contact and all communications to the final transaction would be advantageous. Interrogation of scam lifecycles focusing on the flow of information could pave the way for a strengthened approach in identifying scam processes rather than relying on just static features. From such research more detailed inferences could be made about scammer business process.

The next stage of research for this body of knowledge is the investigation of methods for automation for the processing natural language. The collection and derivation of scam static features is a very time consuming task. A concern raised is investigator bias in the identification of static features. With the aim of accounting for such concerns and speeding up the process of content analysis of scam descriptions, and later, written scamming accounts, a view to automation is expected. While the process of automation is outside the scope of this research, it is thought that a system could be developed which would allow for the input of a body of text – scam description or victim account, and a content analysis would automatically run which would search the input text, comparing the words used against a pre-identified list of target words, sentences and phrases (static features). The automated process would then use the output information which would be a list of present/non-present static features to aggregate the new scam case into its appropriate scam genre, based upon its static feature composition.

The methodology applied here could also be useful if expanded to include all known types of cybercrime and traditional crimes. The content analysis of crime descriptions, definitions, and victim accounts leading to the identification of static features for each crime family would be useful in the identification of business processes for each crime type. This approach could also benefit from the addition of weighted features which would assist in the rigorous categorization of crime types by adding a third dimension to the data – time, sequences such as order of events. Not only would this process assist in the understanding of crime-type architecture, but it would aid towards the development of greater understanding of criminal business processes. The automation of such a system may be useful in the identification of business models and attributing those found to known organized crime groups.

Further to this, this research involves the identification of scam business processes and through the use of common business methodologies such as risk analysis and critical path analysis, Scam Priority Interference Metrics (SPIM) could be produced which could assist investigators in predicting possible paths of active scam perpetrations and transactions based upon limited histories with applied confidence and accuracy.

5.4 Conclusion

The purpose of this research was to form homogeneous groups of scam perpetrations through the use of hierarchical clustering. It was identified that a hierarchy of fraud exists and through logical deduction it was implied that a hierarchy of scams exists. Deriving data from publicly available scam descriptions, hierarchical clustering analysis was used to ensure homogeneous partitioning of scams into similar clusters which were inferred from the scam static features. The result of this procedure was then tested for reliability by a discriminant function analysis. This research concluded with seven analogous scam clusters which can now be used for future research. Further to the clustering of scams by their descriptions is the opportunity presented to authorities to standardise scam descriptions as well as assist in the identification of significant static features which can be used to confidently identify the type of scam a scam is.

This research was composed of three research questions:

1. Which binary linkage method (furthest neighbour, between groups, within groups, and nearest neighbour linkage) and binary distance measure (Jaccard or Simple Matching) best partitions scam descriptions into homogeneous groups?

    a. Which cluster result contains the fewest number of groups with the least number of scam descriptions allocated to each group?

2. Can static features be used to determine scam group membership of scam cases?

    b. How many static features are required to determine scam group membership?

    c. What static features are useful in determining scam group membership?

3. Can a discriminant function analysis be used to predict scam group memberships for the hierarchical cluster solutions with the fewest number of clusters and least number of scam cases in each cluster to determine which solution accurately predicts scam group memberships at least 95% of the time?

The furthest neighbour, Jaccard coefficient model of hierarchical cluster analysis provided the best results for the homogeneous partitioning of scam cases. Both of the final results selected for scam membership labelling were of this combination of linkage and distance measures and the final cluster result containing the fewest number of groups and the least number of scam cases was the 7-cluster model.

Static features can be used to confidently determine scam membership of scam cases and for the 7-cluster model, only 68 of the 82 derived static features are required to accurately determine scam group membership. The most significant static features useful in determining scam memberships were the ‘what the scam offered’, ‘the role of the victim’, ‘the goal of the scammer’, and ‘the method of scam introduction’.

Discriminant function analysis was suitable for predicting scam group memberships for the hierarchical cluster solutions with the fewest number of clusters and least number of scam cases in each cluster to determine which solution accurately predicts scam group memberships at least 95% of the time with the final 7-cluster solution achieving 95% accuracy with all insignificant static features removed.

The results of this research contribute to scam and fraud literature as well as extend on current scam and fraud research methodologies by extending on the applications regularly used in focussed scam research and applying them here in this research. Further to this, the reduction of scam events into homogeneous scam clusters will assist investigative and enforcement agencies by reducing time, money and resources spent on scam case investigations. It is also hoped that the results from this research will lead the way towards a common scam lexicon and enhanced coordination and cooperation in transnational taskforces.

This research is exploratory in nature and is therefore affected by some identified research limitations. The methods selected for data analysis were chosen because they had been successfully used in the past on either similar data types or in a related field to that being investigated here. One of the biggest limitations to this research is the subjectiveness and interpretability of the data during the data identification and data gathering phases. Since this research was manually sourced, coded and collected, confidence can be gained in a single subjective and interpretive view which was stable across the whole data identification and data collection phase. However, this manual process proved time consuming and limiting because the researcher was limited to English only data sources and bound by time. Given more time and assistance from non-English speaking individuals, a larger and more representative, comprehensive sample could be attained. Due to the nature of the data sources belonging to similar, related or same jurisdictions and countries, there is a possibility that scam types were repeated across source platforms. Since scam descriptions are assumed to be authored by the source agency, this has not become an issue in the consideration of data suitability because the purpose of this study is to analyse and compare those scam types and genres across related jurisdictions.

To the knowledge of this researcher, the combination of analyses used in this research on the type of scam static features derived from those publicly available descriptions has not been attempted before. Therefore, this study represents an exploratory study into the usefulness of scam static features in predicting group membership and identifying scam genres. Exploratory analyses are troubled by concerns with validity, reliability and reproducibility which are the reasons for testing the reliability of the hierarchical clustering of scam static features using a discriminant function analysis.

In conclusion, this body of research investigated the current state of research on scams. Presented here is an overview of current scamming statistics as well as a comparison of the methodologies used for each information source. Following this, various academic explorations into scam types was presented and the methodologies applied within such research explored. A gap in knowledge was identified and research objectives and research questions presented to address this. The implications of this research were discussed along with the proposed methodology and applied methods for achieving the goal of the study. The research underwent numerous phases of assessment and the analysis of the results revealed significant contributions to the research of scams. A formal methodological process was defined by the success of these results, each research question was successfully answered and this body of research effectively contributes to the field of scams research.

Bibliography

[1] Abonyi, Janos, Feil, & Balazs. (2007). Cluster analysis for data mining and system identification. ISBN: 9783764379872. Retrieved August 2, 2010 from http://reader.eblib.com.au.ezproxy.ballarat.edu.au/Reader.aspx?p=371549&o=168&u=SFG4MaRjCcmBMqRRMlZ8kw%3d%3d&t=1280724536&h=056D1AD922BE0231D6CA35BB129B763596217132&s=3349136&ut=544&pg=1&r=img&pat=n#

[2] Agresti, A. (2002). *Categorical Data Analysis*. (2nd ed). Canada: John Wiley & Sons Inc., Wiley-Interscience.

[3] Airoldi, E. & Malin, B. (2004). *ScamSlam: An architecture for learning the criminal relations behind scam spam.* School of Computer Science, Carnegie Mellon University, Pittsburgh, USA, Tech. Rep. CMU-ISRI-04-121. Retrieved August 2, 2010 from http://reports-archive.adm.cs.cmu.edu/anon/isri2004/CMU-ISRI-04-121.pdf

[4] Alhaija, E. & Richardson, A. (2003). Growth prediction in class III patients using cluster and discriminant function analysis, *The European Journal of Orthodontics, 25*(6), 599 – 608. Retrieved August 2, 2010 from http://ejo.oxfordjournals.org/cgi/content/short/25/6/599

[5] Anderson, D., Fleizach, C., Savage, S., & Voelker, G. (2007). Spamscatter: Characterizing Internet scam hosting infrastructure, *Usenix Security.* Retrieved August 2, 2010 from http://cseweb.ucsd.edu/~voelker/pubs/spamscatter-security07.pdf

[6] Aranganayagi, S. & Thangavel,K. (2009). Clustering categorical data using the Bayesian concept, *International Journal of Computer Theory and Engineering,* 1(2), 119-125. Retrieved August 2, 2010 from www.ijcte.org/papers/019.pdf

[7] Australian Bureau of Statistics, (2008). *Personal Fraud 2007.* (4528.0). Canberra: Australian Bureau of Statistics. Retrieved August 2, 2010 from http://www.abs.gov.au/ausstats/abs@.nsf/mf/4528.0

[8] Australian Centre for Policing Research & The Australian Transaction Reports and Analysis Centre Proof of Identity Steering Committee. (2006). *Standardisation of definitions of identity crime terms: A step towards consistency* (145.3). Australia: Commonwealth of Australia. Retrieved Feb 12, 2009, from, www.acpr.gov.au.

[9] Australian Competition and Consumer Commission. *ACCC.* Retrieved Sept 20, 2009, from http://www.accc.gov.au/content/index.phtml/itemId/142.

[10] Australian Competition and Consumer Commission. *Scamwatch*. Retrieved Sept 18, 2009, from http://www.scamwatch.gov.au/content/index.phtml/tag/scamwatch/.

[11] Australian Competition and Consumer Commission. (2008). *The Little Black Book of Scams.* ACCC: Canberra, ACT. ISBN: 978 1 921393 22 8.

[12] Australasian Legal Information Institute. *Cybercrime Act 2001 (Cwlth).* Retrieved December 10, 2008, from www.austlii.edu.au

[13] Australian Research Council. (2009). *Strategic Plan.* Canberra: Australian Research Council. Retrieved August 2, 2010 from http://www.arc.gov.au/about_arc/strategic_plan.htm

[14] Berkhin, P. (2002). *Survey of clustering data mining techniques*, UCR California: Accrue Software Inc., Springer. Retrieved August 2, 2010 from www.ee.ucr.edu/~barth/EE242/**clustering_survey**.pdf

[15] Birzer, M. & Craig-Moreland, D. (2008). Using discriminant analysis in policing research, *Professional Issues in Criminal Justice,* 3(2), 33-48.

[16] Calais, P., Pires, D., Guedes, D., Meira, J., Hoepers, C. & Steding-Jessen, K. (2008). *A campaign-based characterization of spamming strategies,* in Proceedings of the 5th Conference on E-mail and Anti-Spam (CEAS). Retrieved August 2, 2010 from http://www.google.com.au/url?sa=t&source=web&cd=1&ved=0CBkQFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.161.5368%26rep%3Drep1%26type%3Dpdf&ei=oCdWTID-O4zEvQPzxrgY&usg=AFQjCNEUi-1-3uFChhiv8255QaPkECeelw&sig2=i53zAsNGYaar_Tvhxj1m_A

[17] Chau, D.H., & Faloutsos, C. (2005). *Fraud detection in electronic auction*, in European Web Mining Forum EWMF/APKDD. Retrieved August 2, 2010 from www.cs.cmu.edu/~dchau/papers/chau_**fraud_detection**.pdf

[18] Choi, K. (2008). Computer crime victimization and integrated theory: an empirical assessment, *International Journal of Cyber Criminology,* 2(1), 308-333. Retrieved August 2, 2010 from http://www.allacademic.com/meta/p_mla_apa_research_citation/1/2/6/7/1/p126710_index.html

[19] Choo, K.R., Smith, R.G., McCusker, R. (2007). *Future directions in technology-enabled Crime: 2007-2009* (no. 78). Canberra, Australian Capital Territory: Australian Institute of Criminology. Retrieved August 2, 2010 from http://www.aic.gov.au/publications/current%20series/rpp/61-80/rpp78.aspx

[20] Choo, K.R. & Smith, R.G. (2008). “Criminal exploitation of online systems by organised crime groups”. *Asian Criminology*, vol. 3, pp. 37-59. Retrieved August 2, 2010 from http://www.springerlink.com/content/l437117571870577/

[21] Creswell, J. & Thomas Jr., L. (2009, January 25). The talented Mr. Madoff, *The New York Times.* Retrieved Jan 13, 2010, from http://www.nytimes.com/2009/01/25/business/25bernie.html.

[22] Denman, K., Ferris, J., Greig, B., Hutchins, S., McGauran, J., Kerr, D. & Thompson, C. (2004). *Cybercrime,* Parliamentary Joint Committee on the Australian Crime Commission, Parliament of the Commonwealth of Australia (ISBN 0 642 71327 8). Canberra: Senate Printing Unit.

[23] Dolan, K. (2004). Internet auction fraud: the silent victims, *Journal of Economic Crime Management,* 2, 1-22. Retrieved August 2, 2010 from https://www.utica.edu/academic/institutes/ecii/publications/articles/BA2DF0D2-D6ED-10C7-9CCB88D5834EC498.pdf

[24] Drummond, A. (2010, January 20). *Skimming Fraud Costing Aussies Millions,* *http://au.biz.yahoo.com/100120/2/2atkk.html*.

[25] Dyrud, M. (2005). *I brought you a good news: An analysis of Nigerian 419 letters*, in Proceedings of the Association for Business Communication, Annual Convention. Retrieved August 2, 2010 from http://www.businesscommunication.org/conventionsNew/proceedingsNew/2005New/PDFs/07ABC05.pdf

[26] Eisenhardt, K. (1989). Building theories from case study research, *Academy of Management Review, Academy of Management,* 14(4), 532-550. Retrieved August 2, 2010 from http://pages.cpsc.ucalgary.ca/~sillito/cpsc-601.23/readings/eisenhardt-1989.pdf

[27] Environics Research Group. (2008). *2007 Canadian Consumer Mass Marketing Fraud Survey*, (459-06), Ontario CA: Environics Research Group. Retrieved August 2, 2010 from http://www.ic.gc.ca/eic/site/ic1.nsf/vwapj/Environics-Competition%20Bureau-MMF-FinalRReport-Feb2008.pdf/$file/Environics-Competition%20Bureau-MMF-FinalRReport-Feb2008.pdf

[28] Federal Bureau of Investigation. *Common Fraud Scams.* Retrieved Oct 12, 2009, from http://www.fbi.gov/majcases/fraud/fraudscams.htm.

[29] Fernández, W. (2004). *The Grounded Theory Method and Case Study Data in IS Research: Issues and Design,* The Australian National University: ANU E-Press, 43-59. Retrieved August 2, 2010 from http://epress.anu.edu.au/info_systems/part-ch05.pdf

[30] FIDO. (2010). *Australian Securities and Investments Commission Financial Tips and Safety Checks, Scams and Warnings.* Retrieved Sept 21, 2009, from http://www.fido.gov.au/fido/fido.nsf/byHeadline/Scams%20%26%20Swindlers%20portal.

[31] Francetič, M., Nagode, M. & Nastav, B. (2005). Hierarchical clustering with concave data sets, *Metodoloski zvezki,* 2(2), 173-193. Retrieved August 2, 2010 from http://mrvar.fdv.uni-lj.si/pub/mz/mz2.1/francetic.pdf

[32] Glickman, H. (2005). The Nigerian '419' advance fee scams: prank or peril? *Canadian Journal of African Studies/Revue Canadienne des Études Africaines, Canadian Association of African Studies*, 39, 460-489. Retrieved August 2, 2010 from www.jstor.org/stable/25067495

[33] Goode, M., Musgrave, P., Byrne, G., Tannin, G., Perks, N., Hardy, F., Mayo, N., Alderson, K., Sturgess, V., Cochrane, S., Thomas, L. & Cairns, L. (2008). *Final Report: Identity Crime,* Model Criminal Law Officers' Committee of the Standing Committee of Attorneys-General (ISBN 1 921241 37 3). ACT: Commonwealth of Australia. Retrieved August 2, 2010 from http://www.ag.gov.au/www/agd/rwpattach.nsf/VAP/%28CFD7369FCAE9B8F32F341DBE097801FF%29~6Final+Report+Identity+Crime+March+2008.PDF/$file/6Final+Report+Identity+Crime+March+2008.PDF

[34] Hays, H. & Prenzler, T. (2002). *Profiling fraudsters: A Queensland case study in fraudster crime, final report to Crime Prevention Queensland.* Queensland, Australia: Griffith

University, Department of Criminology and Criminal Justice. Retrieved August 2, 2010 from http://aic.gov.au/en/crime_types/economic/fraud/~/media/aic/research/fraud/profilingfraudsters.ashx

[35] Hennig, C. (2007). Cluster-wise assessment of cluster stability, *Computational Statistics and Data Analysis, Elsevier,* 52, 258-271. Retrieved August 2, 2010, from www.ucl.ac.uk/stats/research/reports/psfiles/rr271.pdf

[36] Holt, T. & Graves, D. (2007). Qualitative analysis of advance fee fraud e-mail scams, *International Journal of Cyber Criminology,* 1(1), 137-154.

[37] Internet Crime Complaint Center. (2008). *2008 Internet Crime Report.* Retrieved Oct, 1, 2009, from *http://www.ic3.gov/media/annualreport/2008_IC3Report.pd.*

[38] Internet Crime Complaint Center. (2007). *2007 Internet Crime Report.* Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2007_IC3Report.pdf.*

[39] Internet Crime Complaint Center. (2006). *Internet Crime Report*. Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2006_IC3Report.pdf.*

[40] Internet Crime Complaint Center. (2005). *IC3 2005 Internet Crime Report*. Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2005_IC3Report.pdf.*

[41] Internet Crime Complaint Center. (2004). *IC3 2004 Internet Fraud Crime Report*. Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2004_IC3Report.pdf.*

[42] Internet Crime Complaint Center. (2003). *IC3 2003 Internet Fraud Report.* Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2003_IC3Report.pd.*

[43] Internet Fraud Complaint Center. (2002). *IFCC 2002 Internet Fraud Report.* Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2002_IC3Report.psf.*

[44] Internet Fraud Complaint Center. (2001). *IFCC 2001 Internet Fraud Report*. Retrieved Oct 1, 2009, from *http://www.ic3.gov/media/annualreport/2001_IC3Report.psf.*

[45] Internet Crime Complaint Center. *Internet Crime Scams,* Retrieved Oct 1, 2009, from http://www.ic3.gov/crimescams.aspx.

[46] Jagatic, T., Johnson, N., Jakobsson, M. & Menczer, F. (2007). Social phishing, *Communications of the ACM, 50*(10), 94-100. Retrieved August 2, 2010 from http://portal.acm.org/citation.cfm?id=1290968

[47] Jie, J., Wang, G., Qin, Y. & Chau, M. (2004). Crime data mining: A general framework and some examples, *IEEE Computer, Citeseer, 37*, 50-56. Retrieved August 2, 2010 from http://www.computer.org/portal/web/csdl/doi/10.1109/MC.2004.1297301

[48] Lamp, J., Milton, S. & Melbourne, V. (2007). *Indexing research: An approach to grounding Ingarden's ontological framework information systems foundations: theory, representation and reality,* Australia: ANU E-Press. Retrieved August 2, 2010 from http://lamp.infosys.deakin.edu.au/pubs/06_isf_operatontol.pdf

[49] Lea, S., Fischer, P. & Evans, K. (2009). *The Psychology of Scams: Provoking and Committing Errors of Judgement*., United Kingdom: *Office of Fair Trading: University of Exeter School of Psychology,* OFT1070. Retrieved August 2, 2010, from http://www.oft.gov.uk/shared_oft/reports/consumer_protection/oft1070.pdf

[50] Looks Too Good To Be True. *Types of Fraud,* Retrieved Oct 12, 2009, from http://www.lookstoogoodtobetrue.com/fraud.aspx.

[51] Lösch, A. (2006). Combining quantitative methods and grounded theory for researching e-reverse auctions, *Libri, 56*, 133-144. Retrieved August 2, 2010, from www.librijournal.org/pdf/2006-3pp133-144.pdf

[52] Mauldin, T. (2008). *Discriminant analysis: pathological gambling*, JMP Users, 24. Retrieved August 2, 2010, from www.jmp.com/about/newsletters/jmpercable/pdf/24_spring_2008.pdf

[53] Meyer, A., Garcia, A., Souza, A. & Souza, Jr. S. (2004). Comparison of similarity coefficients used for cluster analysis with dominant markers in maize, *Genetics and Molecular Biology, 27*(1), 83-91. Retrieved August 2, 2010, from www.cababstractsplus.org/abstracts/Abstract.aspx?AcNo=20043072993

[54] Moore, T. & Clayton, R. (2007). *An empirical analysis of the current state of phishing attack and defence*, in Workshop on the Economics of Information Security. Retrieved August 2, 2010 from http://weis2007.econinfosec.org/papers/51.pdf

[55] Office of Fair Trading. (2006). *Research on Impact of Mass Marketed Scams*, (OFT833), United Kingdom: Office of Fair Trading. Retrieved August 2, 2010, from http://www.oft.gov.uk/shared_oft/reports/consumer_protection/oft883.pdf

[56] On Guard Online. *Email Scams,* Retrieved Oct 5, 2009, from http://www.onguardonline.gov/topics/email-scams.aspx.

[57] Pyryt, M. (2004). Pegnato revisited: Using discriminant analysis to identify gifted children, *Psychology Science,* 46(3), 342-347. Retrieved August 2, 2010, from www.pabst-publishers.de/psychology-science/3-2004/07.pdf

[58] Queensland Police Service. *Scams and frauds.* Retrieved Sept 25, 2010, from http://www.police.qld.gov.au/services/qpol/.

[59] Rowlands, B. (2005). Grounded in practice: Using interpretive research to build theory, *Electronic Journal of Business Research Methods, 3*(1), 81-92. Retrieved August 2, 2010, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.134.9432

[60] Scam Smart. *Scam Smart awareness and Prevention – Scams, Spams and Malware.,* Retrieved Sept, 24, 2009, from http://www.scamsmart.com.au/prevention/index.php.

[61] Stabek, A. Brown, S. & Watters, P. (2009). *The Case for a Consistent Cyberscam Classification Framework (CCCF)*, in Proceedings of the Cybercrime and Trustworthy Computing Workshop,

2009. Pp 525-530. Retrieved August 2, 2010, from http://portal.acm.org/citation.cfm?id=1638449

[62] United States Computer Emergency Response Team. *US-Cert*. Retrieved June 25, 2009, from http://www.us-cert.gov/.

[63] United States Department of Justice & Federal Bureau of Investigation. (2010, January 11). *Unlicensed Orange County Mortgage Broker Sentenced to Nearly Six Years in Prison in $40 Million Fraud Scam,* Press Release. Retrieved Jan 13, 2010, from http://losangeles.fbi.gov/dojpressrel/pressrel10/la011110.htm.

[64] United States Department of Justice & Federal Bureau of Investigation. (2010, January 11). *Hedge Fund Manager Who Bilked Relatives Out of $25 Million Sentenced to Over 10 Years in Federal Prison Press Release, Press* Release. Retrieved Jan 13, 2010, from http://losangeles.fbi.gov/dojpressrel/pressrel10/la011110a.htm.

[65] United States History. *Charles Ponzi.* Retrieved on May 28, 2009 from http://www.u-s-history.com

[66] United States Department of Justice & Federal Bureau of Investigation. (2010, January 12). *Former Webster Bank Employee, Husband Arrested: Couple Charged with Defrauding of Approximately $6.2 Million,* Press Release. Retrieved Jan 13, 2010, from http://newhaven.fbi.gov/dojpressrel/pressrel10/nh01210a.htm

[67] United States Department of Justice & Federal Bureau of Investigation. (2010, January 13). *Purchasing Official at a New York City Hospital Pleads Guilty to Bid Rigging,* Press Release. Retrieved Jan 12, 2010, from http://newyork.fbi.gov/dojpressrel/pressrel10/nyfo011210a.htm.

[68] United States Department of Justice & Federal Bureau of Investigation. (2010, January 13). *Leader of $47 Million Mortgage Fraud Scam Sentenced to Prison: Mortgage Fraud Scam Used Web of Companies and False Loan Documents*, Press Release. Retrieved Jan 13, 2010, from http://seattle.fbi.gov/dojpressrel/pressrel10/se010810b.htm.

[69] United States Department of Justice & Federal Bureau of Investigation. (2010, January 11). *Defendant Pleads Guilty to Telemarketing Stock Fraud Scam: Reece Operated Boiler Rooms in Marietta; Targeted Foreign Victim Investors.* Press Release. Retrieved Jan 13, 2010, from http://atlanta.fbi.gov/dojpressrel/pressrel10/atl011110.htm.

[70] United States Postal Inspectors Service. *Mail Fraud Scams*, Retrieved June 4, 2009, from https://postalinspectors.uspis.gov/investigations/MailFraud/fraudscams/FraudScams.aspx.

[71] Wahlert, G. (1998). *Crime in cyberspace: Trends in computer crime in Australia,* Internet Crime Conference, Melbourne, Australia. Retrieved August 2, 2010, from http://www.afp.gov.au/media-centre/publications/platypus/previous-editions/1998/june-1998/cyber.aspx

[72] Weaver, R. & Collins, M.P. (2007). *Fishing for phishies: Applying capture-recapture methods to estimate phishing populations*, in Anti-Phishing Working Group eCrime Researcher

Summit. Retrieved August 2, 2010, from [www.cert.org/netsa/.../ecrimes07-collins-weaver-fish-for-**phish**.pdf](www.cert.org/netsa/.../ecrimes07-collins-weaver-fish-for-phish.pdf)

[73] Witten, I.H. & Frank, E., D.D. Cerra (ed.). (2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, USA: *Morgan Kaufmann Publishers.*

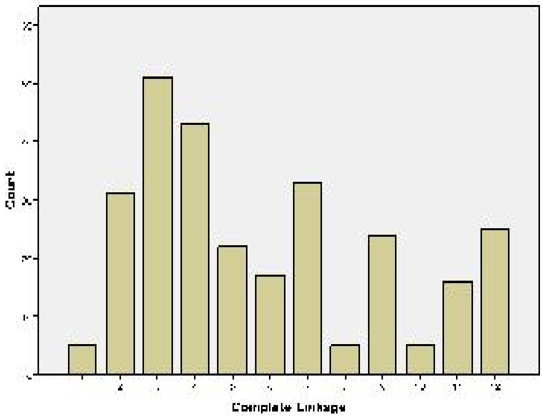
Glossary

Cyberscam	A scam committed through the use of Internet technology
Cyberscammer	One who commits a cyberscam
Fraud	Acquisition of something of value through deceptive means
Fraudster	One who commits fraud
Homogenous	Relatively equal, similar, even
Scam	A type of fraud, a tool used to acquire something of value through deceptive means
Scammer	One who commits a scam
Semantic	Human focused
Static feature	A single stable element of a larger group of elements
Syntactic	Machine or technology focused
Technology based crime	A crime that relies on the use of technology at some stage throughout its lifecycle
Technology enabled crime	A crime that requires the use of technology
Technology enhanced crime	A crime that does not require but is enhanced by the use of technology
Transnational	Multinational

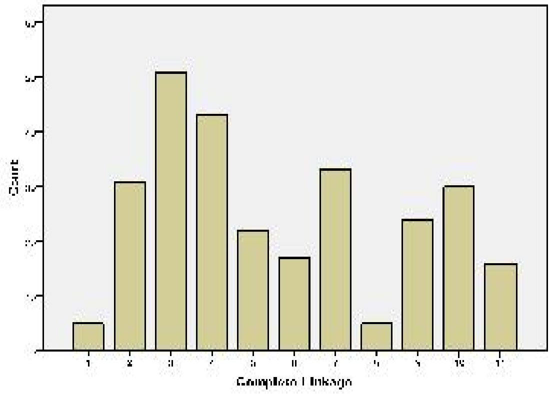
Appendix

Table 40: Scam Static Features

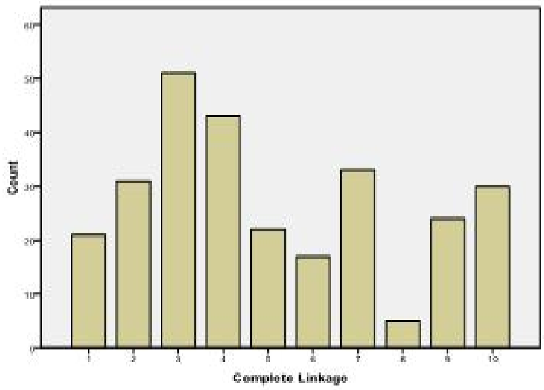
Features	Type	Features	Type
Seller	Role of the victim	Love affection and connection	What the scheme claimed
Customer	Role of the victim	Government agency	What the scheme claimed
Target Specific	Role of the victim	Large return	What the scheme claimed
Unassociated	Role of the victim	Effective	What the scheme claimed
Received	Method of introduction	Refund available	What the scheme claimed
Introduced	Method of introduction	Fraudulent activity	What the scheme claimed
Sought	Method of introduction	No credit check required	What the scheme claimed
Website	Tool for scheme proliferation	Quick response	What the scheme required from the victim
Face to face	Tool for scheme proliferation	Confidentiality	What the scheme required from the victim
Text message	Tool for scheme proliferation	Payment of upfront costs	What the scheme required from the victim
Phone call	Tool for scheme proliferation	Receive and send funds	What the scheme required from the victim
Seminar	Tool for scheme proliferation	Call a premium number	What the scheme required from the victim
Internet forum	Tool for scheme proliferation	Transfer excess	What the scheme required from the victim
Internet pop up	Tool for scheme proliferation	Complete sale outside of auction	What the scheme required from the victim
Email	Tool for scheme proliferation	Send onto others	What the scheme required from the victim
Post	Tool for scheme proliferation	Recruit others	What the scheme required from the victim
Advertisement	Tool for scheme proliferation	Supply personal information	What the scheme required from the victim
Fax	Tool for scheme proliferation	Supply bank account information	What the scheme required from the victim
Prize or money	What the scheme offered	Investment	What the scheme required from the victim
Human interaction	What the scheme offered	Make a donation	What the scheme required from the victim
Financial return	What the scheme offered	Use alternative shipment	What the scheme required from the victim
Membership	What the scheme offered	Syntactic	Method of the scheme
Advice or assistance	What the scheme offered	Semantic	Method of the scheme
Overpayment	What the scheme offered	Compromised website or phony website	Scammers toolbox
Treatment	What the scheme offered	Disguised as invoice	Scammers toolbox
Employment	What the scheme offered	Inferior merchandise	Scammers toolbox
Opportunity for self or others	What the scheme offered	Use of flasified forms	Scammers toolbox
Holiday	What the scheme offered	Use of paraphernalia	Scammers toolbox
Financial services	What the scheme offered	Goods never sent	Scammers toolbox
Good luck	What the scheme offered	Story based	Scammers toolbox
Property	What the scheme offered	Verifiable street address	Scammers toolbox
Share tips	What the scheme offered	Looks genuine	Scammers toolbox
Services	What the scheme offered	Exploitation of legitimate business	Scammers toolbox
Merchandise	What the scheme offered	Testimonials	Scammers toolbox
Partial payment	What the scheme offered	Reward greater than upfront cost	Scammers toolbox
Insight	What the scheme claimed	Further contact by email or phone	Scammers toolbox
Legal	What the scheme claimed	Polite broken English	Scammers toolbox
From financial institution	What the scheme claimed	Financial gain	Goal of the scheme
Information update required	What the scheme claimed	Information gathering	Goal of the scheme
Government approved	What the scheme claimed	Participation	Goal of the scheme



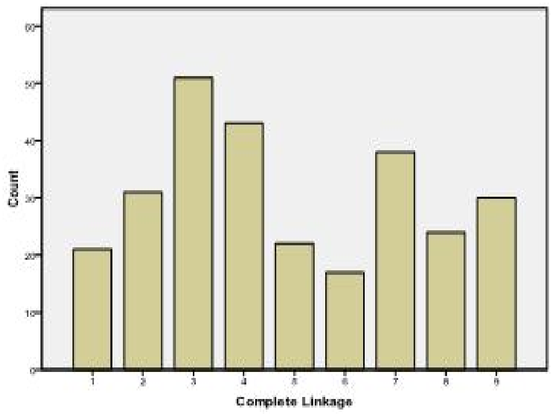
A) 12 Cluster Solution – Cluster Membership Frequency



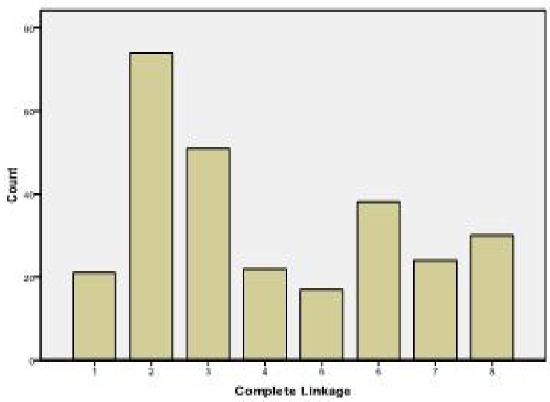
B) 11 Cluster Solution - Cluster Membership Frequency



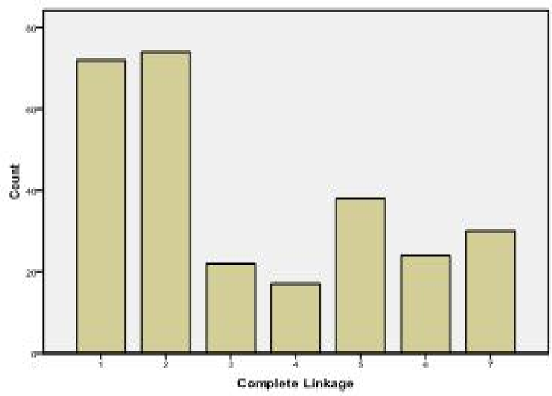
C) 10 Cluster Solution – Cluster Membership Frequency



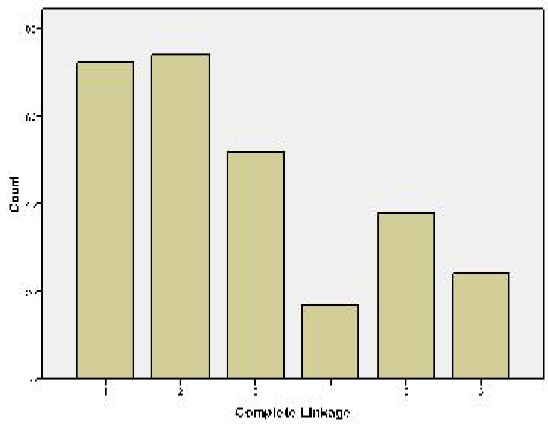
D) 9 Cluster Solution - Cluster Membership Frequency



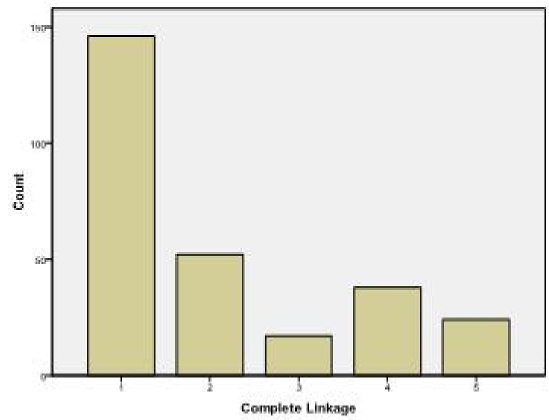
E) 8 Cluster Solution - Cluster Membership Frequency



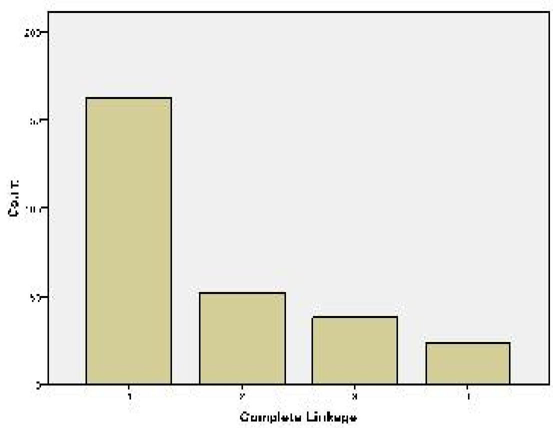
F) 7 Cluster Solution - Cluster Membership Frequency



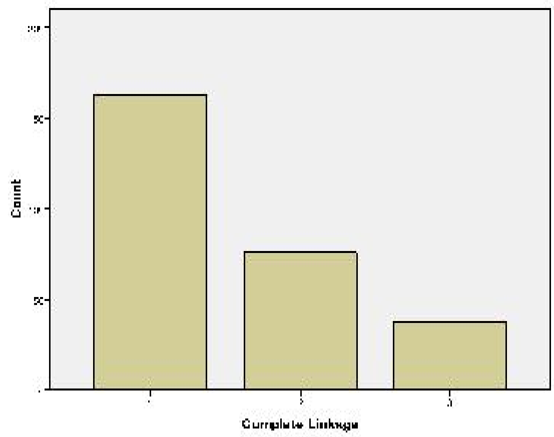
G) 6 Cluster Solution - Cluster Membership Frequency



H) 5 Cluster Solution - Cluster Membership Frequency



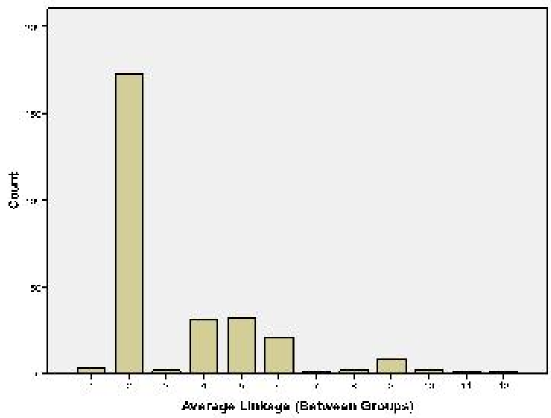
I) 4 Cluster Solution - Cluster Membership Frequency



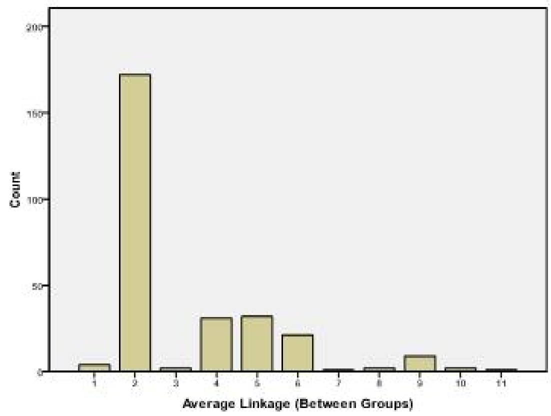
J) 3 Cluster Solution - Cluster Membership Frequency

Figure 14: HCA Furthest Neighbour Jaccard Coefficient Cluster Membership Frequencies[3]

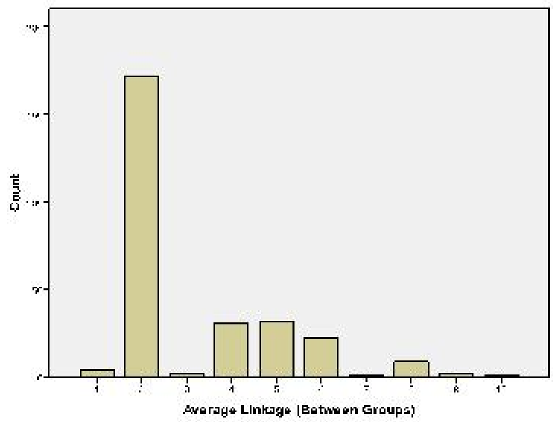
[3] The x-axis is the number of clusters created, the y-axis is the cluster frequency



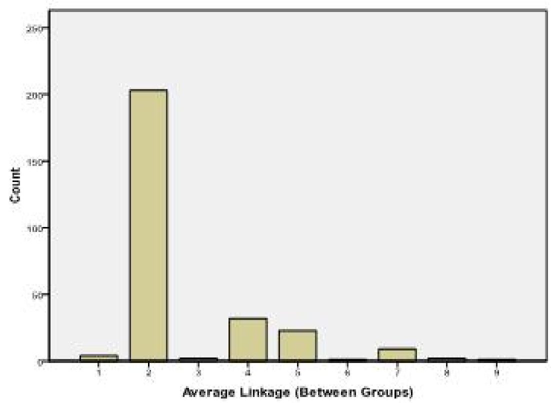
A) 12 Cluster Solution – Cluster Membership Frequency



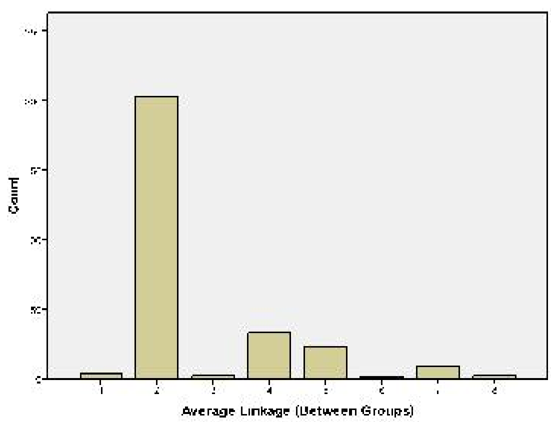
B) 11 Cluster Solution – Cluster Membership Frequency



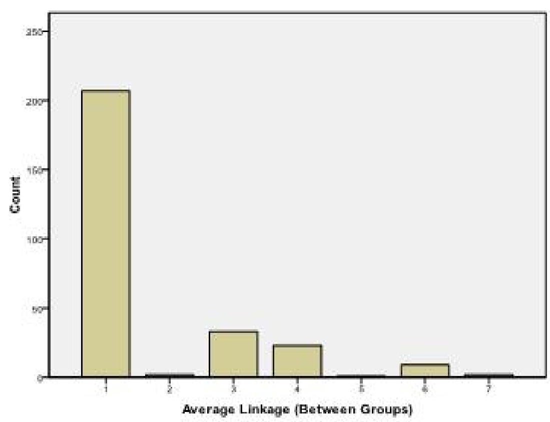
C) 10 Cluster Solution – Cluster Membership Frequency



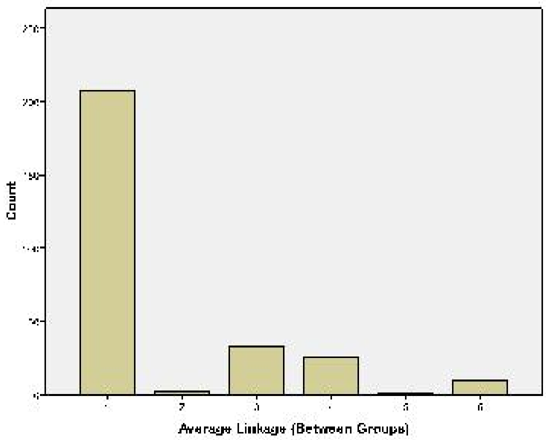
D) 9 Cluster Solution – Cluster Membership Frequency



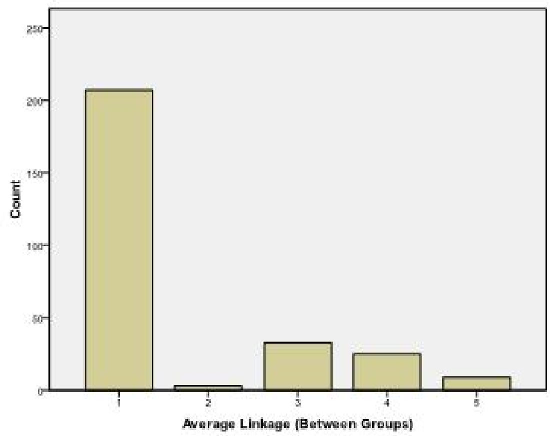
E) 8 Cluster Solution – Cluster Membership Frequency



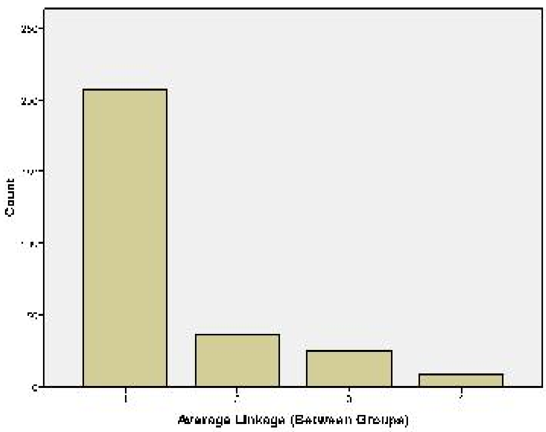
F) 7 Cluster Solution – Cluster Membership Frequency



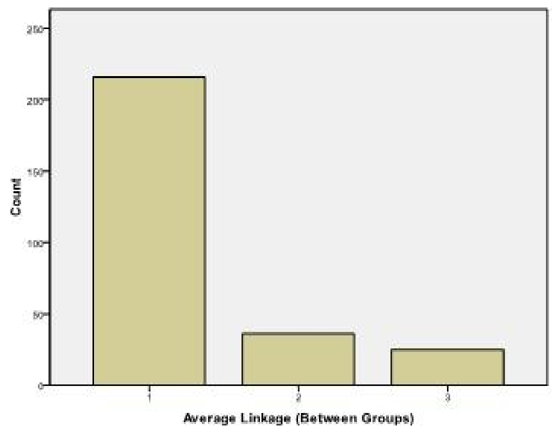
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency

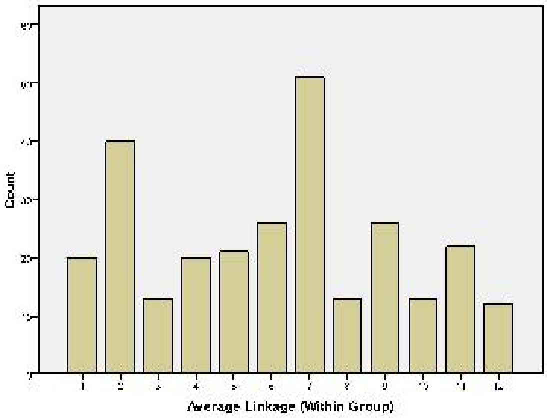


I) 4 Cluster Solution – Cluster Membership Frequency

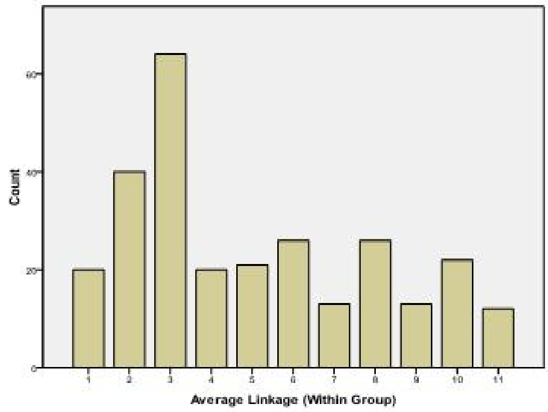


J) 3 Cluster Solution – Cluster Membership Frequency

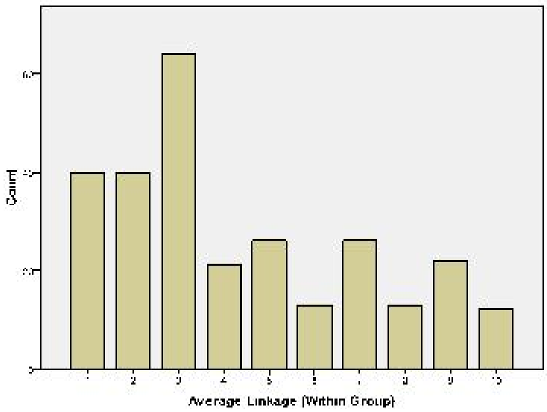
Figure 15: HCA Between Groups Linkage Jaccard Coefficient Cluster Membership Frequencies



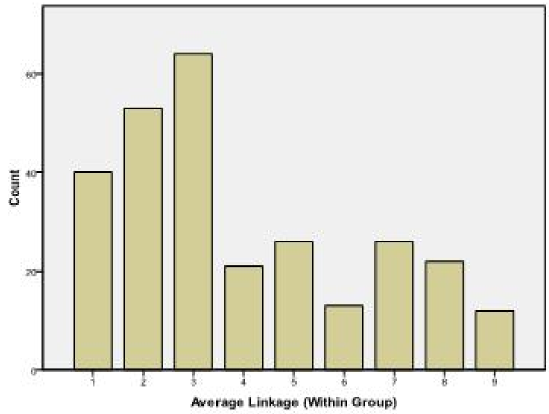
A) 12 Cluster Solution – Cluster Membership Frequency



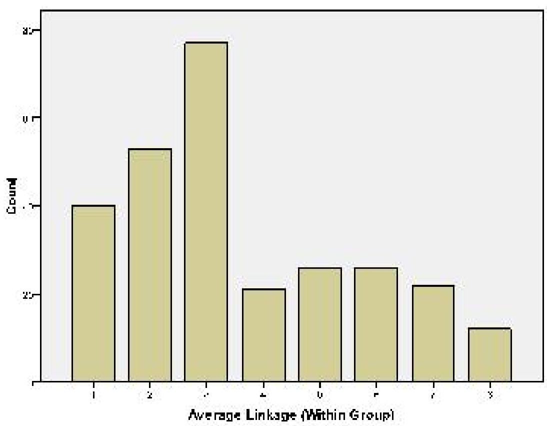
B) 11 Cluster Solution – Cluster Membership Frequency



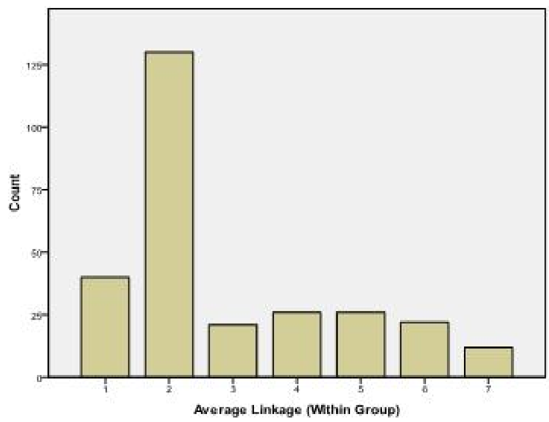
C) 10 Cluster Solution – Cluster Membership Frequency



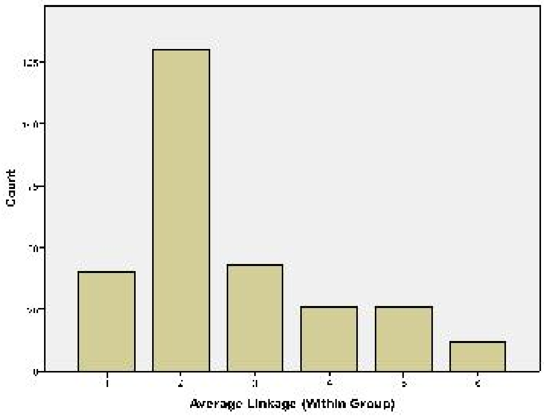
D) 9 Cluster Solution – Cluster Membership Frequency



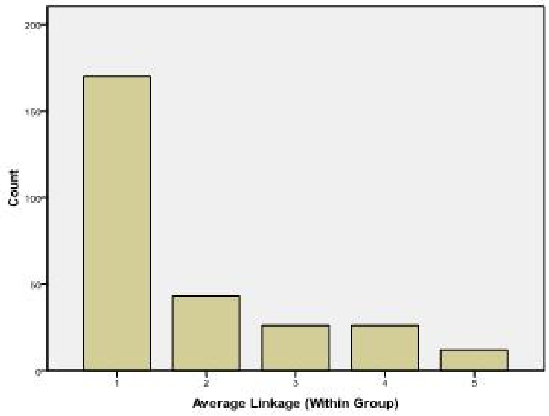
E) 8 Cluster Solution – Cluster Membership Frequency



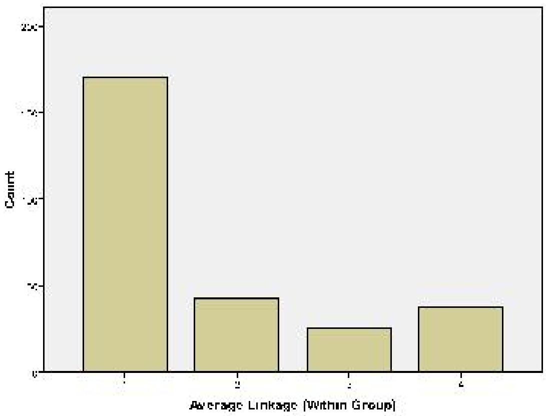
F) 7 Cluster Solution – Cluster Membership Frequency



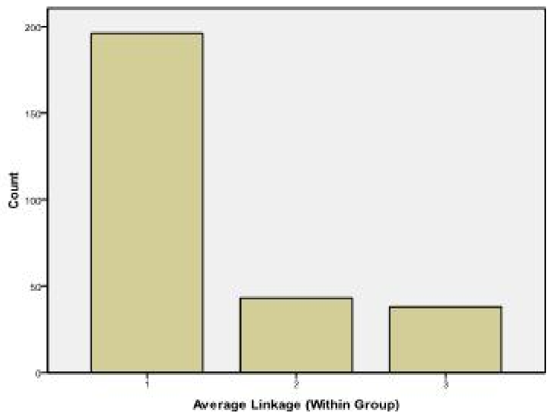
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency

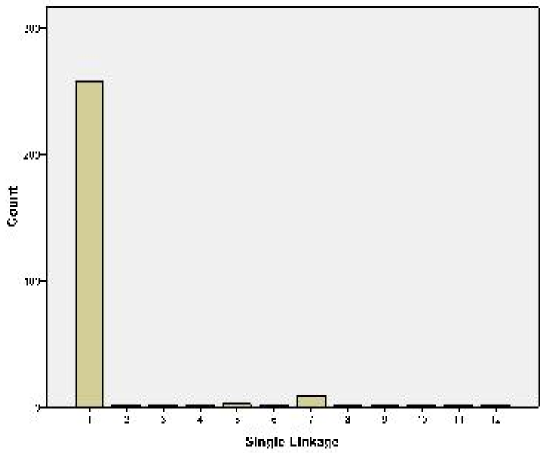


I) 4 Cluster Solution – Cluster Membership Frequency

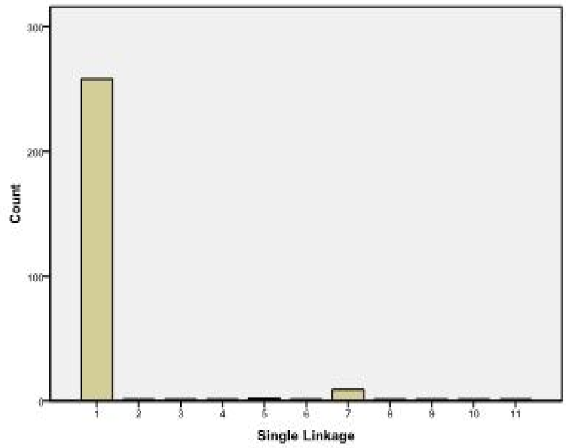


J) 3 Cluster Solution – Cluster Membership Frequency

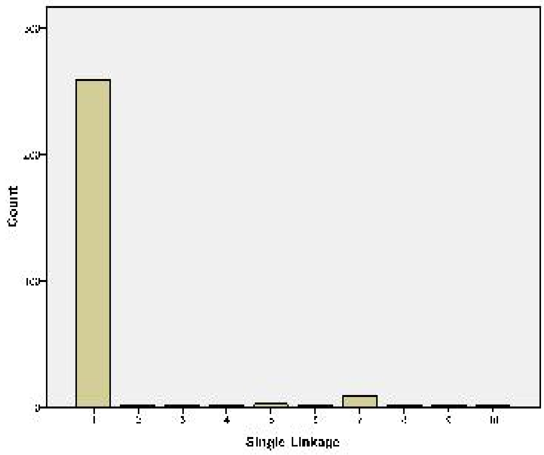
Figure 16: HCA Within Groups Linkage Jaccard Coefficient Cluster Membership Frequencies



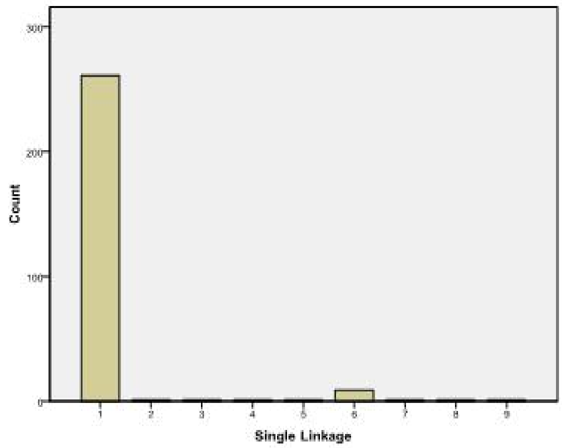
A) 12 Cluster Solution – Cluster Membership Frequency



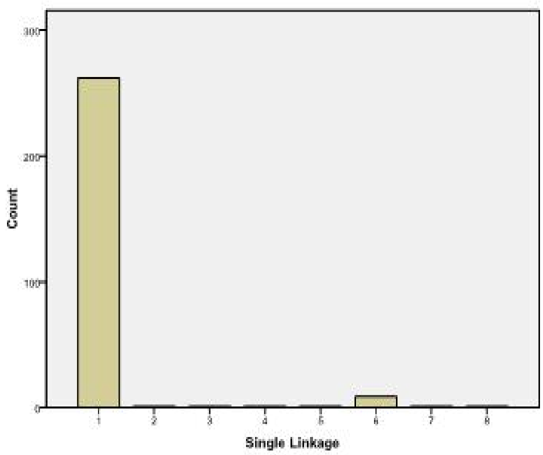
B) 11 Cluster Solution – Cluster Membership Frequency



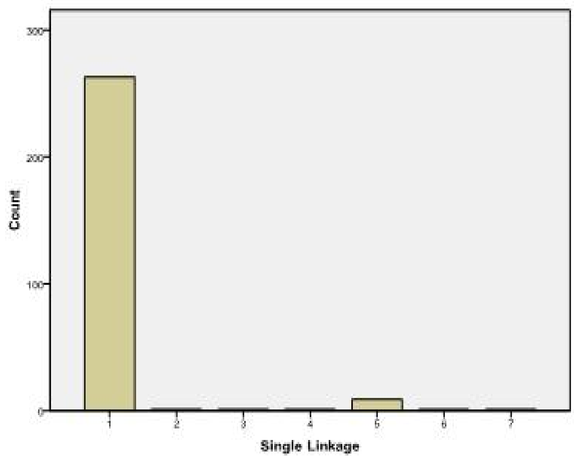
C) 10 Cluster Solution – Cluster Membership Frequency



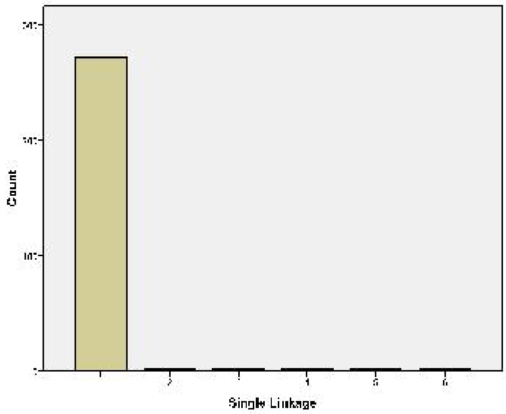
D) 9 Cluster Solution – Cluster Membership Frequency



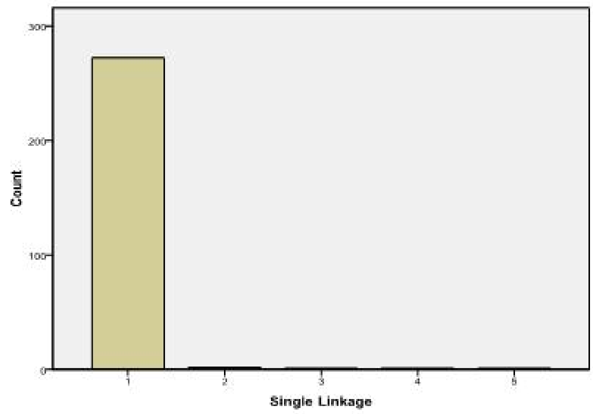
E) 8 Cluster Solution – Cluster Solution Frequency



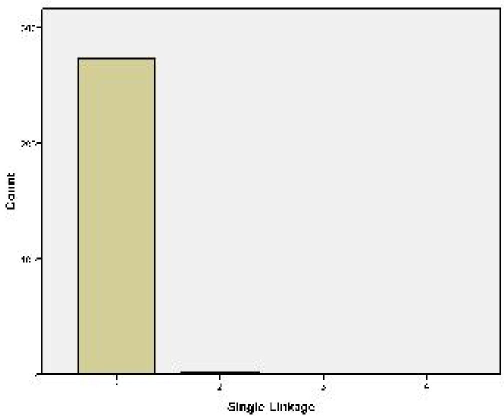
F) 7 Cluster Solution – Cluster Solution Frequency



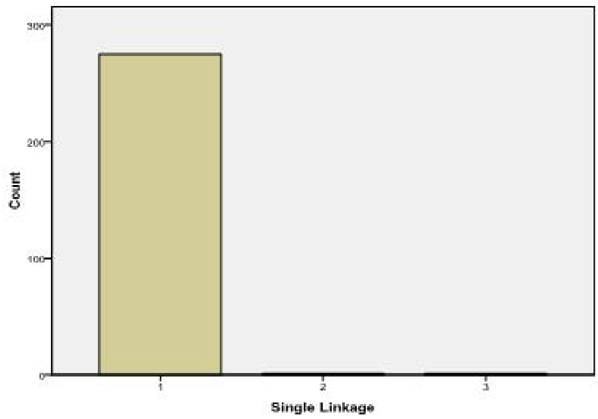
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency

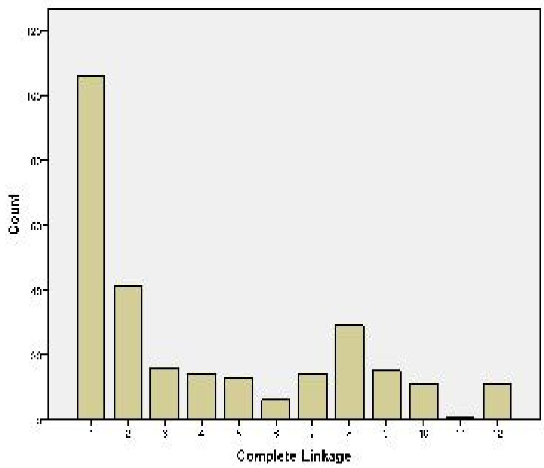


I) 4 Cluster Solution – Cluster Membership Frequency

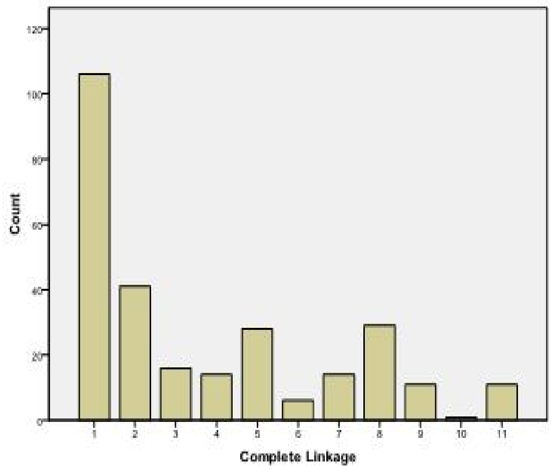


J) 3 Cluster Solution – Cluster Membership Frequency

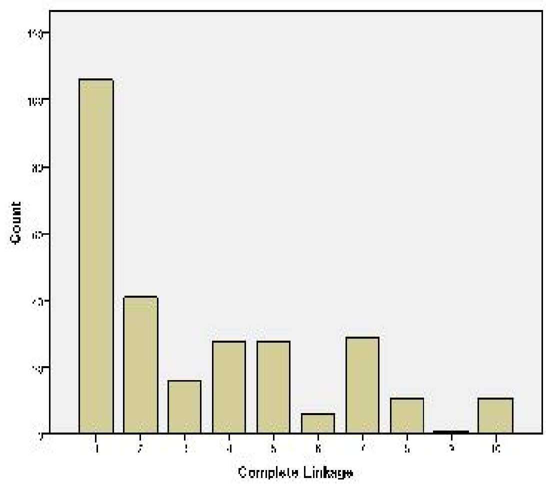
Figure 17: HCA Nearest Neighbour Jaccard Coefficient Cluster Membership Frequencies



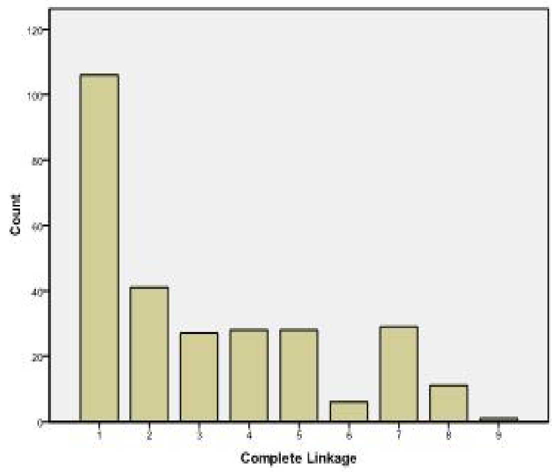
A) 12 Cluster Solution – Cluster Membership Frequency



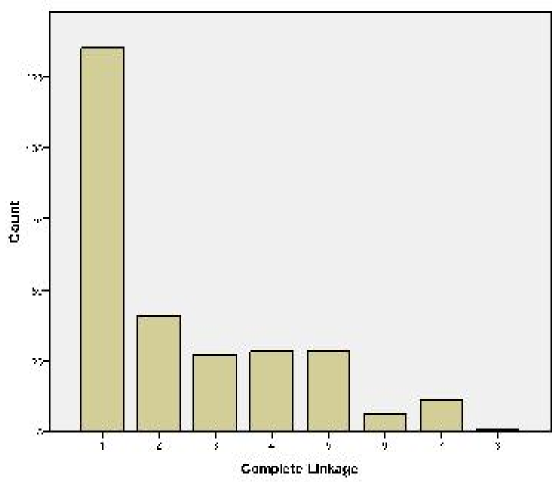
B) 11 Cluster Solution – Cluster Membership Frequency



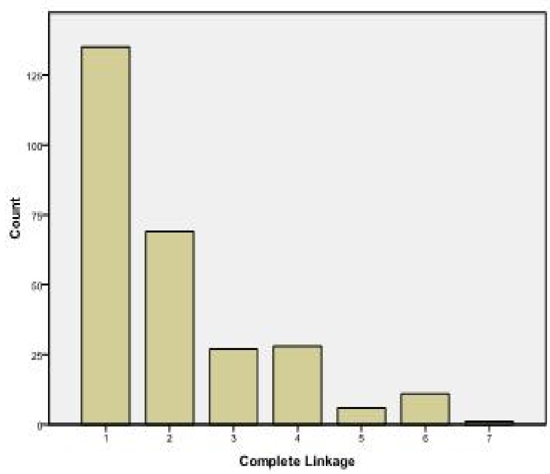
C) 10 Cluster Solution – Cluster Membership Frequency



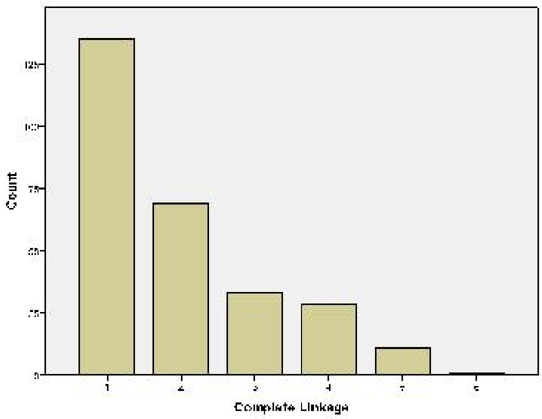
D) 9 Cluster Solution – Cluster Membership Frequency



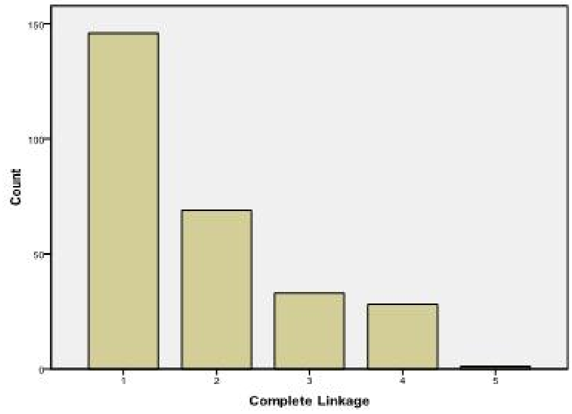
E) 8 Cluster Solution – Cluster Membership Frequency



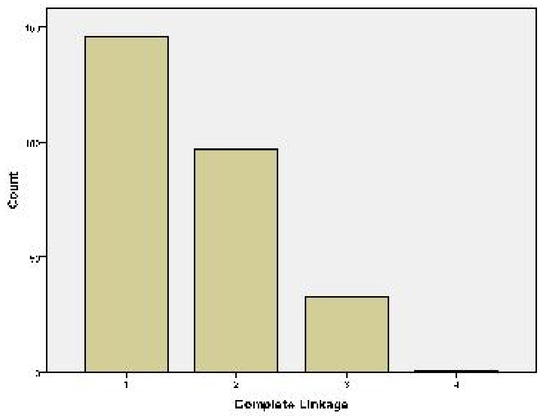
F) 7 Cluster Solution – Cluster Membership Frequency



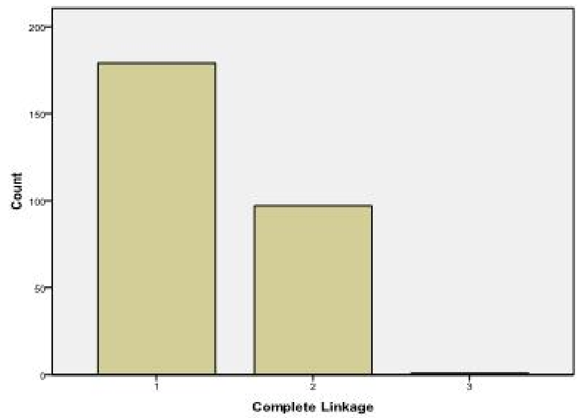
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency

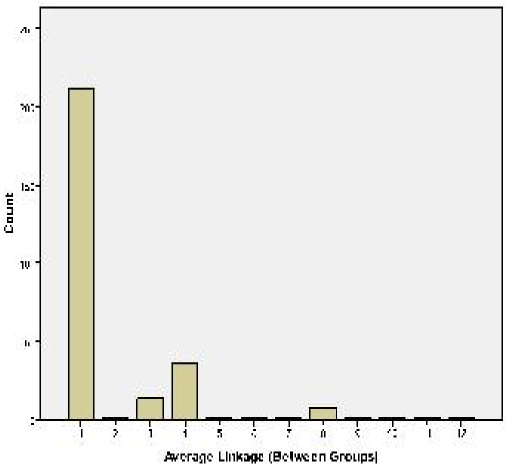


I) 4 Cluster Solution – Cluster Membership Frequency

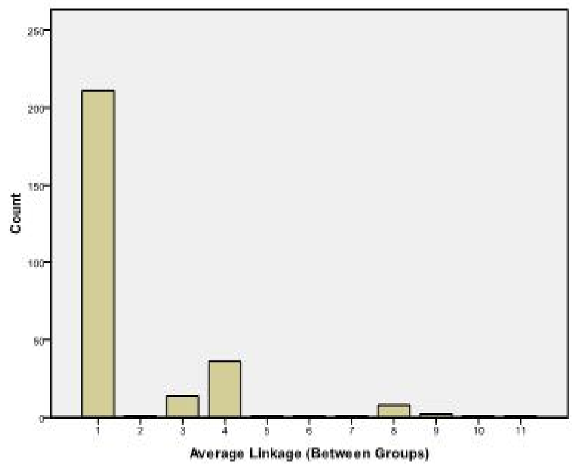


J) 3 Cluster Solution – Cluster Membership Frequency

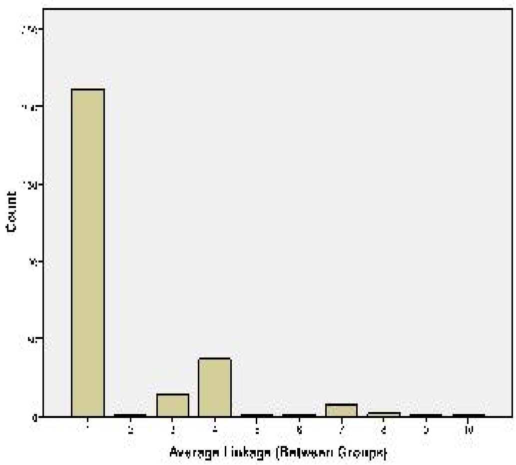
Figure 18: HCA Furthest Neighbour Simple Matching Coefficient Cluster Membership Frequencies



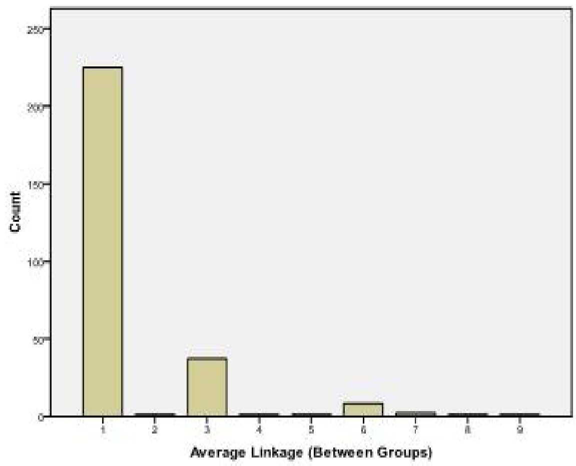
A) 12 Cluster Solution – Cluster Membership Frequency



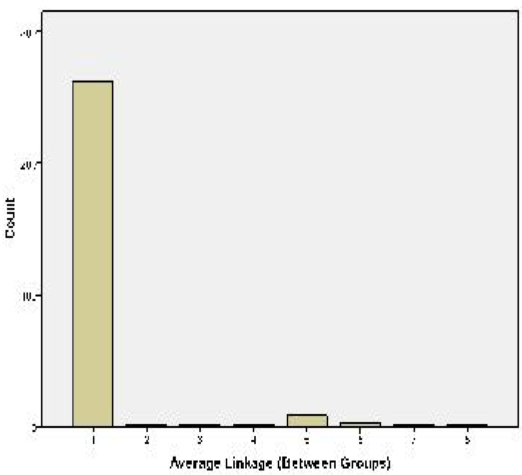
B) 11 Cluster Solution – Cluster Membership Frequency



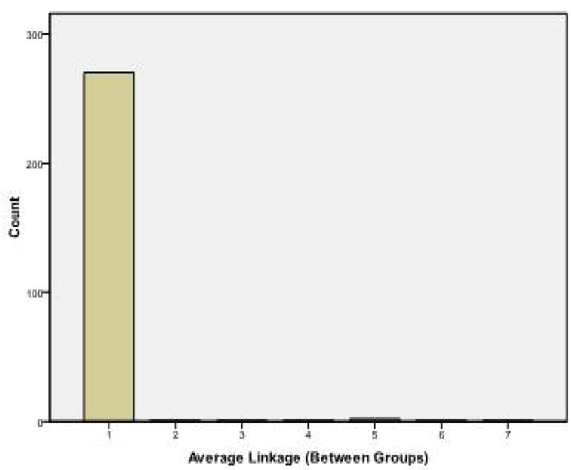
C) 10 Cluster Solution – Cluster Membership Frequency



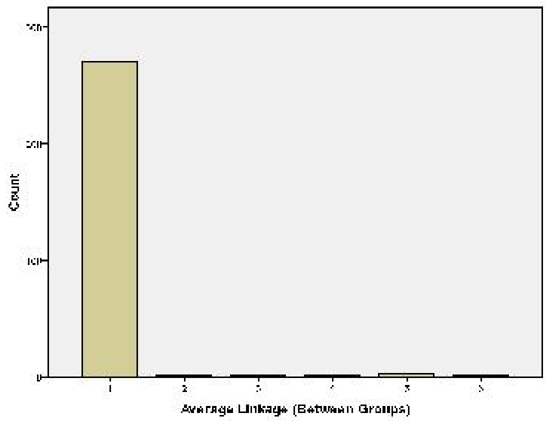
D) 9 Cluster Solution – Cluster Membership Frequency



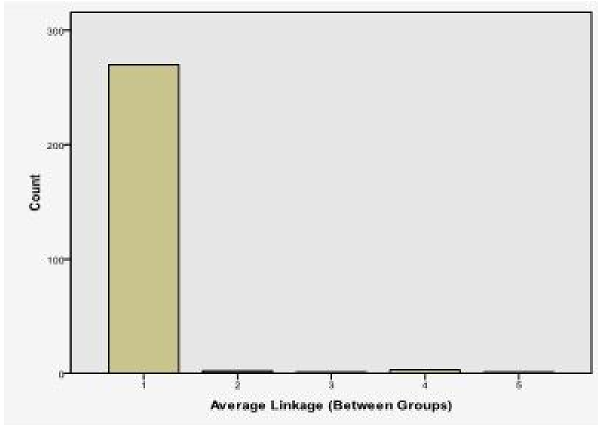
E) 8 Cluster Solution – Cluster Membership Frequency



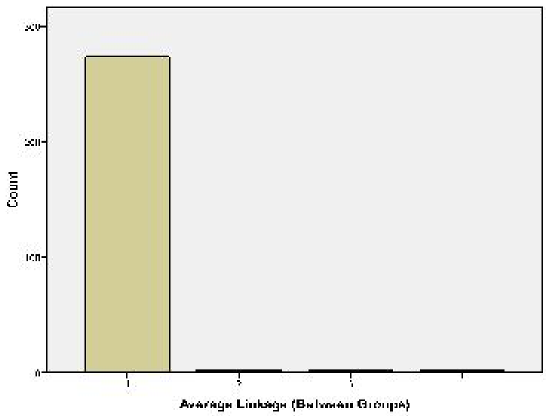
F) 7 Cluster Solution – Cluster Membership Frequency



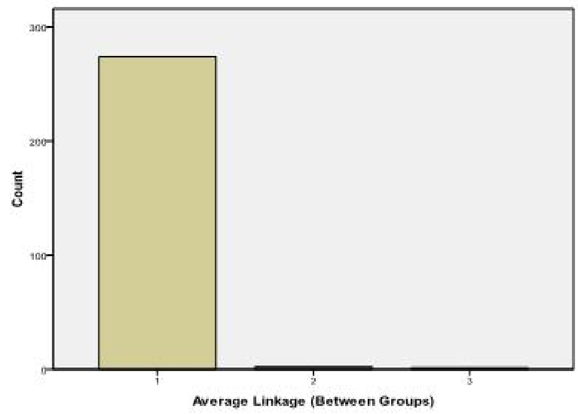
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency

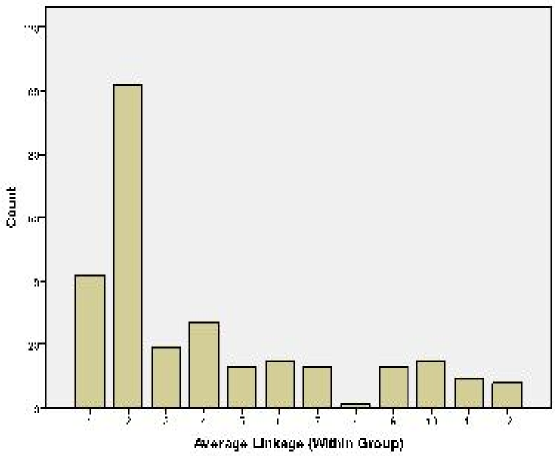


I) 4 Cluster Solution – Cluster Membership Frequency

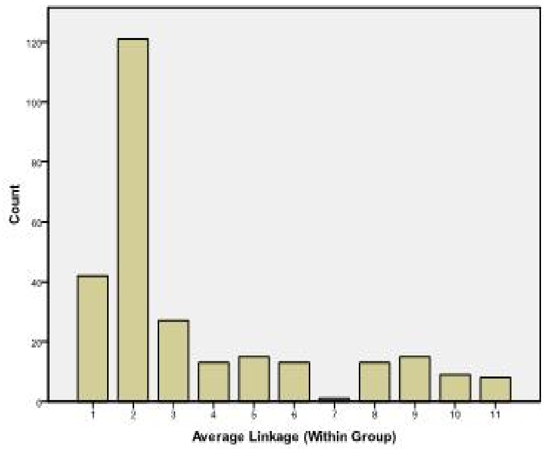


J) 3 Cluster Solution – Cluster Membership Frequency

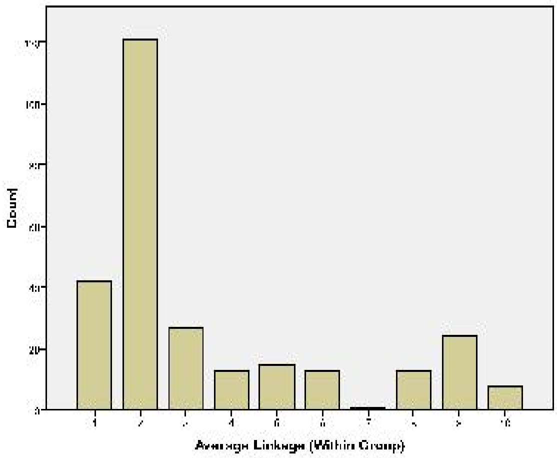
Figure 19: HCA Between Groups Linkage Simple Matching Coefficient Cluster Membership Frequencies



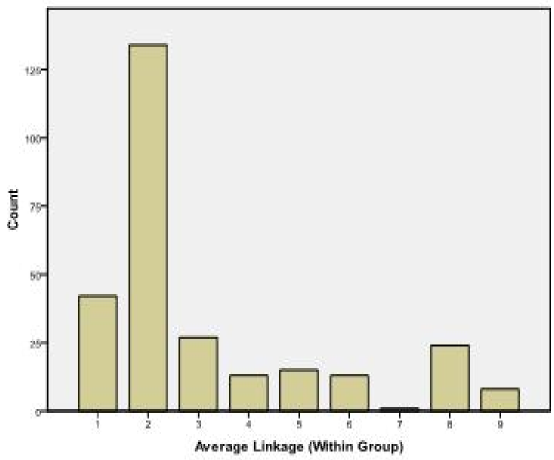
A) 12 Cluster Solution – Cluster Membership Frequency



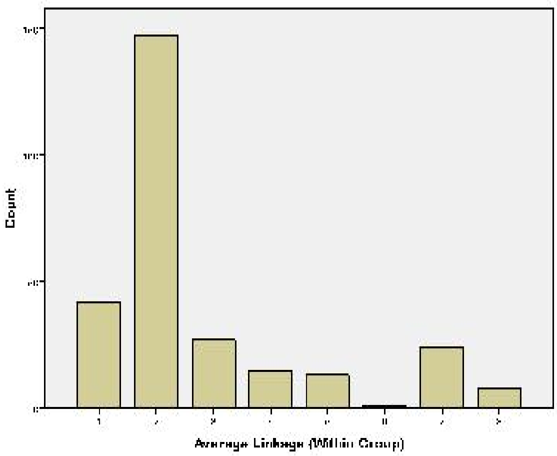
B) 11 Cluster Solution – Cluster Membership Frequency



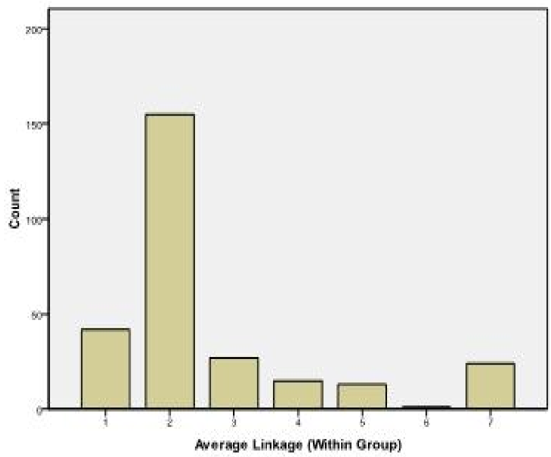
C) 10 Cluster Solution – Cluster Membership Frequency



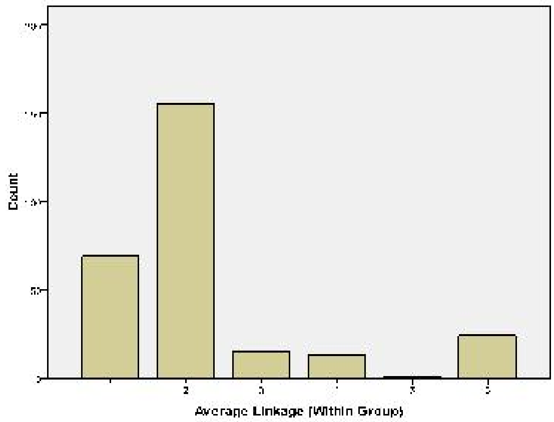
D) 9 Cluster Solution – Cluster Membership Frequency



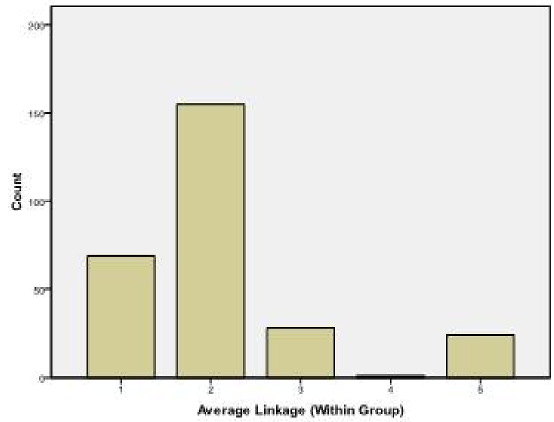
E) 8 Cluster Solution – Cluster Membership Frequency



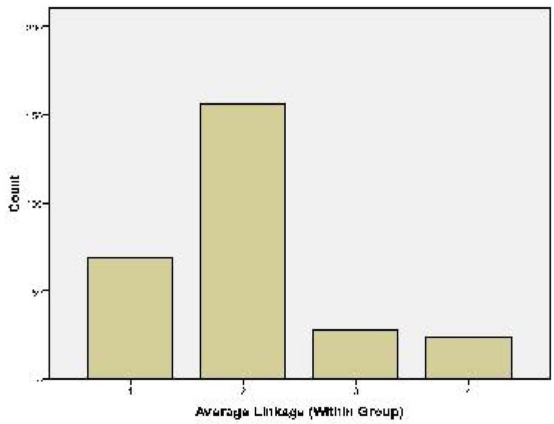
F) 7 Cluster Solution – Cluster Membership Frequency



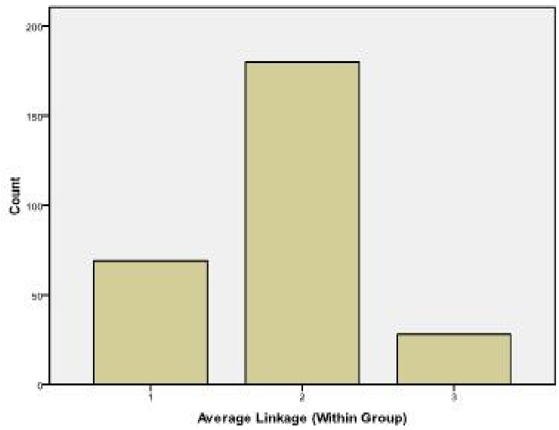
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency

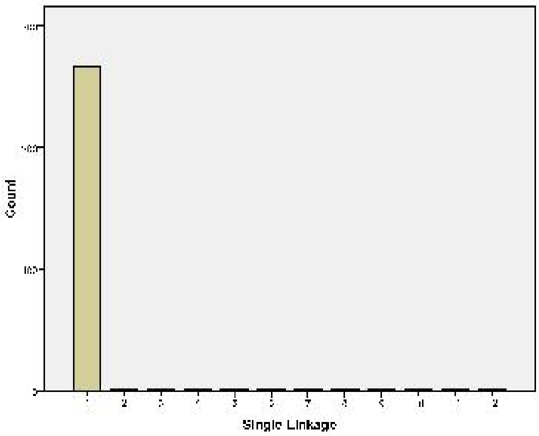


I) 4 Cluster Solution – Cluster Membership Frequency

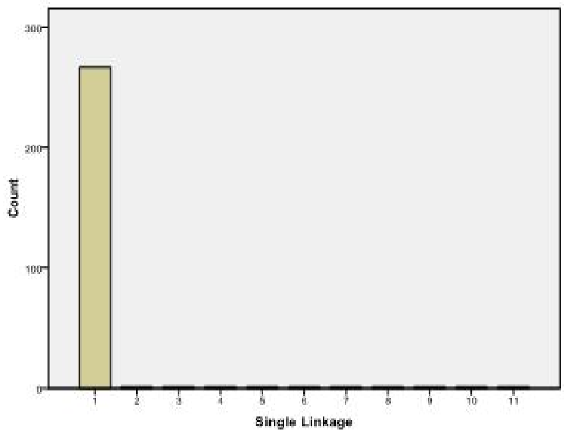


J) 3 Cluster Solution – Cluster Membership Frequency

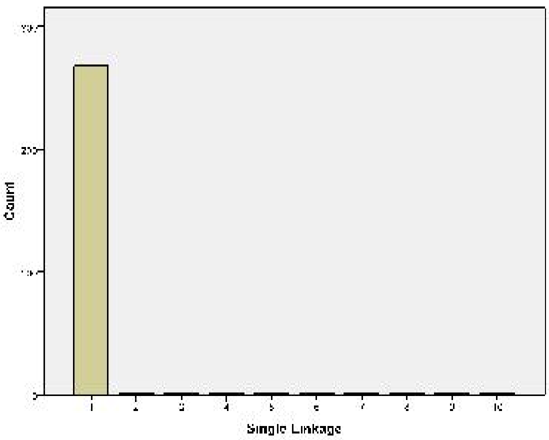
Figure 20: HCA Within Groups Linkage Simple Matching Coefficient Cluster Membership Frequencies



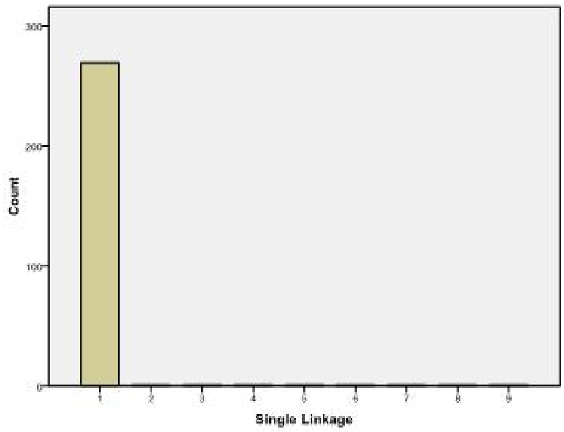
A) 12 Cluster Solution – Cluster Membership Frequency



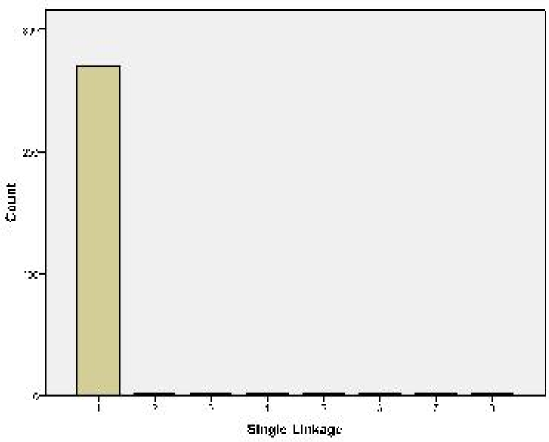
B) 11 Cluster Solution – Cluster Membership Frequency



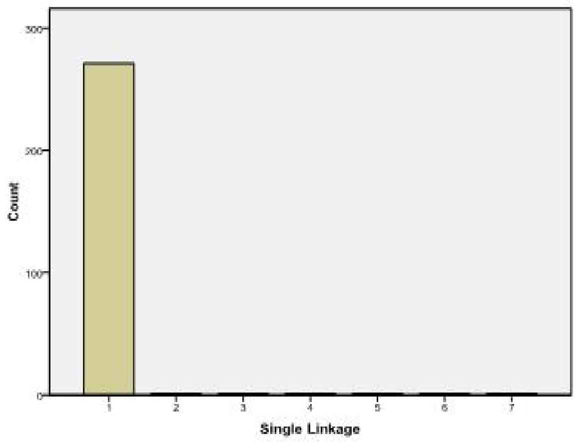
C) 10 Cluster Solution – Cluster Membership Frequency



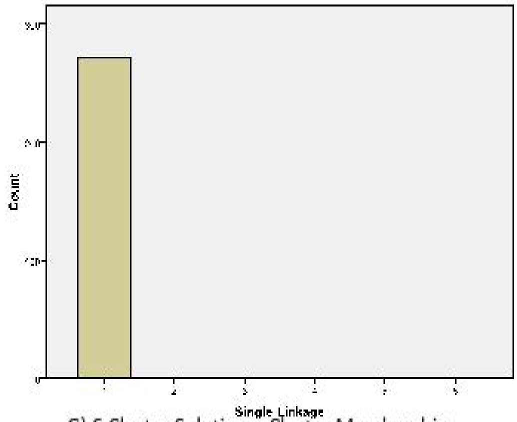
D) 9 Cluster Solution – Cluster Membership Frequency



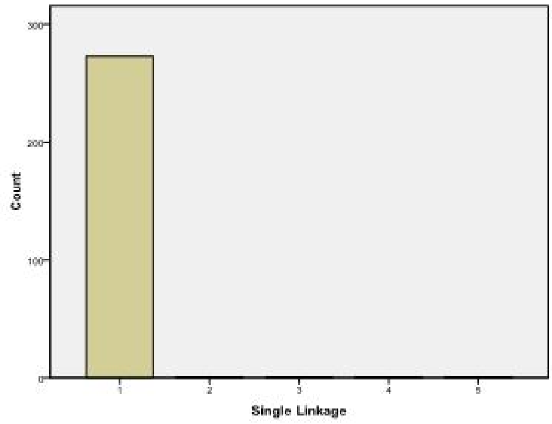
E) 8 Cluster Solution – Cluster Membership Frequency



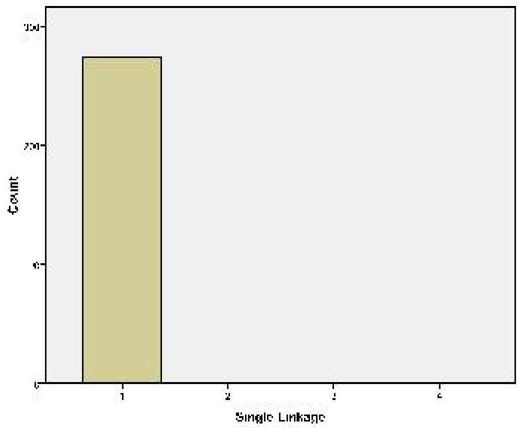
F) 7 Cluster Solution – Cluster Membership Frequency



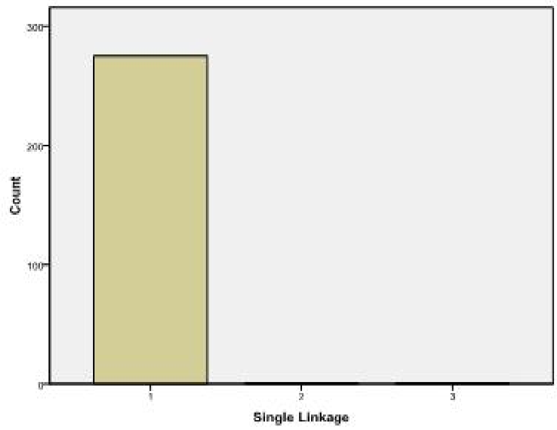
G) 6 Cluster Solution – Cluster Membership Frequency



H) 5 Cluster Solution – Cluster Membership Frequency



I) 4 Cluster Solution – Cluster Membership Frequency



J) 3 Cluster Solution – Cluster Membership Frequency

Figure 21: HCA Nearest Neighbour Jaccard Coefficient Cluster Membership Frequencies

Table 41: DFA 9 Cluster Results Tests of Equality of Group Means for the HCA Furthest Neighbour Jaccard Coefficient Model

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.591	23.220	8	268	.000
Customer	.267	92.086	8	268	.000
TargetSpecific	.938	2.208	8	268	.027
Unassociated	.224	115.835	8	268	.000
Received	.523	30.586	8	268	.000
Introduced	.775	9.698	8	268	.000
Sought	.648	18.218	8	268	.000
WebsiteorOnlineAuction	.800	8.392	8	268	.000
Face2Face	.833	6.734	8	268	.000
Text	.820	7.359	8	268	.000
Phone	.863	5.323	8	268	.000
Seminar	.909	3.354	8	268	.001
InternetForum	.828	6.977	8	268	.000
InternetPopUp	.788	9.004	8	268	.000
Email	.704	14.111	8	268	.000
Post	.746	11.421	8	268	.000
Advertisement	.742	11.663	8	268	.000
Fax	.920	2.904	8	268	.004
PrizeorMoney	.421	46.152	8	268	.000
HumanInteraction	.888	4.219	8	268	.000
FinancialReturn	.564	25.859	8	268	.000
Membership	.857	5.575	8	268	.000
AdviceorAssistance	.877	4.711	8	268	.000
Overpayment	.591	23.220	8	268	.000
Treatment	.888	4.241	8	268	.000
Employment	.156	181.338	8	268	.000
OpportunityForSelfOrOthers	.444	41.874	8	268	.000
Holiday	.856	5.654	8	268	.000
FinancialServices	.879	4.593	8	268	.000
GoodLuck	.970	1.043	8	268	.404
Property	.941	2.089	8	268	.037
Services	.913	3.197	8	268	.002
Merchandise	.655	17.637	8	268	.000
PartialPayment	.923	2.782	8	268	.006
Insight	.958	1.476	8	268	.166
Legal	.895	3.919	8	268	.000
FromFinancialInstitution	.867	5.123	8	268	.000
DetailUpdateorConfirmationRequired	.709	13.776	8	268	.000
GovernmentApproved	.940	2.120	8	268	.034
LoveAffectionConnection	.879	4.608	8	268	.000
GovernmentAgency	.963	1.269	8	268	.260
LargeReturn	.602	22.102	8	268	.000
Effective	.871	4.944	8	268	.000
RefundAvailable	.971	.997	8	268	.439
FraudulentActivity	.882	4.474	8	268	.000
ShareTips	.921	2.885	8	268	.004
NoCreditCheckRequired	.977	.800	8	268	.603
LittleorNoRisk	.852	5.824	8	268	.000
FromCorporateOrGovOfficial	.861	5.400	8	268	.000
QuickResponse	.937	2.236	8	268	.025
Confidentiality	.834	6.670	8	268	.000
PayupFrontCosts	.735	12.055	8	268	.000
ReceiveAndSendFunds	.719	13.085	8	268	.000
CallaPremiumNumber	.716	13.280	8	268	.000
TransferExcess	.633	19.458	8	268	.000
CompleteSaleoutsideofAuction	.923	2.782	8	268	.006
SendOntoOthers	.919	2.968	8	268	.003
RecruitOthers	.735	12.096	8	268	.000
SupplyPersonalInformation	.762	10.491	8	268	.000
SupplyBankAccDetails	.779	9.503	8	268	.000
Invest	.681	15.708	8	268	.000
MakeADonation	.790	8.909	8	268	.000
AlternativeShipment	.767	10.199	8	268	.000
Syntactic	.657	17.472	8	268	.000
Semantic	.680	15.796	8	268	.000
CompromisedWebsiteorFalseWebsite	.790	8.906	8	268	.000
DisguisedasInvoice	.946	1.912	8	268	.058
InferiorMerchandise	.881	4.529	8	268	.000
UseofFalsifiedForms	.847	6.061	8	268	.000
UseofParaphernalia	.822	7.274	8	268	.000
GoodsNeverSent	.803	8.208	8	268	.000
StoryBased	.543	28.168	8	268	.000
VerifiableStreetAddress	.984	.547	8	268	.821
LooksGenuine	.844	6.171	8	268	.000
ExploitLegitBusiness	.914	3.141	8	268	.002
Testimonials	.903	3.610	8	268	.001
RewardGreaterThanUpfrontCosts	.923	2.783	8	268	.006
FurtherContactbyEmailorPhone	.966	1.179	8	268	.312
PoliteBrokenEnglish	.948	1.820	8	268	.073
FinancialGain	.325	69.555	8	268	.000
Information	.483	35.917	8	268	.000
Participation	.650	17.999	8	268	.000

Table 42: DFA 9 Cluster Results Variable Failing Tolerance Testing for the HCA Furthest Neighbour Jaccard Coefficient Model

Variables Failing Tolerance Test[a]

	Within-Groups Variance	Tolerance	Minimum Tolerance
Overpayment	.019	.000	.000

Table 43: DFA 9 Cluster Results Eigenvalues for the HCA Furthest Neighbour Jaccard Coefficient Model

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	14.997[a]	32.8	32.8	.968
2	10.042[a]	22.0	54.8	.954
3	6.590[a]	14.4	69.3	.932
4	4.723[a]	10.3	79.6	.908
5	3.458[a]	7.6	87.2	.881
6	2.352[a]	5.2	92.4	.838
7	1.788[a]	3.9	96.3	.801
8	1.703[a]	3.7	100.0	.794

Table 44: DFA 9 Cluster Results Function Significance Tests for the HCA Furthest Neighbour Jaccard Coefficient Model

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 8	.000	3157.613	648	.000
2 through 8	.000	2517.194	560	.000
3 through 8	.000	1962.410	474	.000
4 through 8	.002	1494.226	390	.000
5 through 8	.009	1091.241	308	.000
6 through 8	.040	745.976	228	.000
7 through 8	.133	466.577	150	.000
8	.370	229.712	74	.000

Table 45: DFA 9 Cluster Results Predicted Groups Memberships for the HCA Furthest Neighbour Jaccard Coefficient Model

		Highest Group					Second Highest Group		
			P(D>d | G=g)						
Case Number	Actual Group	Predicted Group	p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid
1	1	1	.292	8	1.000	9.633	9	.000	60.338
2	2	2	.993	8	1.000	1.463	9	.000	72.784
3	2	2	.338	8	1.000	9.048	9	.000	96.518
4	3	3	.367	8	1.000	8.717	1	.000	37.177
5	3	3	.003	8	1.000	23.276	1	.000	63.210
6	4	4	.281	8	1.000	9.771	6	.000	48.024
7	3	3	.489	8	1.000	7.448	9	.000	40.245
8	5	5	.076	8	1.000	14.244	9	.000	114.789
9	5	5	.908	8	1.000	3.386	4	.000	136.632
10	5	5	.965	8	1.000	2.426	3	.000	160.453
11	3	3	.210	8	1.000	10.863	6	.000	47.693
12	3	3	.957	8	1.000	2.605	9	.000	27.095
13	3	3	.589	8	1.000	6.525	9	.000	47.764
14	6	6	.059	8	1.000	15.007	4	.000	75.830
15	6	6	.137	8	1.000	12.329	3	.000	86.642
16	6	6	.778	8	1.000	4.808	3	.000	52.735
17	7	7	.992	8	1.000	1.537	4	.000	69.158
18	6	6	.154	8	1.000	11.935	7	.000	47.207
19	7	7	.076	8	.986	14.229	3	.008	23.831
20	4	4	.374	8	1.000	8.635	3	.000	50.763
21	3	3	.942	8	1.000	2.876	4	.000	33.914
22	8	8	.785	8	1.000	4.740	1	.000	150.207
23	9	9	.185	8	1.000	11.299	3	.000	68.197
24	8	8	.025	8	1.000	17.536	1	.000	149.131
25	7	7	.000	8	.863	28.587	1	.137	32.262
26	7	7	.629	8	1.000	6.161	6	.000	86.037
27	7	7	.618	8	1.000	6.258	6	.000	86.803
28	1	1	.787	8	1.000	4.721	3	.000	72.620
29	5	5	.846	8	1.000	4.125	7	.000	142.576
30	4	4	.327	8	1.000	9.185	3	.000	49.467
31	2	2	.842	8	1.000	4.168	4	.000	47.665
32	6	6	.070	8	.957	14.486	4	.043	20.704
33	9	9	.609	8	.998	6.338	3	.002	18.322
34	9	9	.943	8	1.000	2.851	3	.000	43.696
35	9	9	.961	8	1.000	2.515	3	.000	19.966
36	4	4	.218	8	1.000	10.721	1	.000	66.675
37	4	4	.965	8	1.000	2.438	6	.000	34.343
38	4	4	.696	8	1.000	5.565	3	.000	36.177
39	9	9	.248	8	1.000	10.250	3	.000	56.196
40	3	3	.969	8	1.000	2.346	9	.000	26.452
41	8	8	.952	8	1.000	2.699	1	.000	132.168
42	8	8	.370	8	1.000	8.679	1	.000	197.078
43	8	8	.962	8	1.000	2.497	1	.000	158.932
44	1	1	.433	8	1.000	8.009	4	.000	95.361
45	7	7	.034	8	.998	16.674	6	.001	29.958
46	7	7	.002	8	.831	24.770	4	.169	27.961
47	8	8	.847	8	1.000	4.108	1	.000	133.711
48	5	5	.997	8	1.000	1.125	3	.000	170.750
49	8	8	1.000	8	1.000	.702	1	.000	134.491
50	7	7	.000	8	.806	50.591	5	.194	53.436

51	1	1	.191	8	1.000	11.192	7	.000	70.626
52	9	9	.193	8	1.000	11.147	4	.000	37.114
53	4	4	.996	8	1.000	1.304	6	.000	34.939
54	2	2	.364	8	1.000	8.748	3	.000	51.980
55	7	7	.752	8	1.000	5.051	3	.000	63.661
56	9	9	.498	8	.789	7.363	3	.211	10.000
57	5	5	.527	8	1.000	7.086	7	.000	150.475
58	7	7	.997	8	1.000	1.204	3	.000	67.100
59	5	5	.515	8	1.000	7.198	6	.000	161.650
60	2	2	.136	8	1.000	12.357	3	.000	28.471
61	4	4	.912	8	1.000	3.336	3	.000	37.976
62	3	3	.636	8	.982	6.104	9	.018	14.060
63	7	7	.731	8	1.000	5.246	3	.000	68.707
64	3	3	.150	8	.756	12.021	9	.244	14.282
65	4	4	.997	8	1.000	1.159	3	.000	26.392
66	3	3	.768	8	1.000	4.904	4	.000	31.906
67	7	7	.172	8	1.000	11.558	3	.000	54.262
68	2	2	.016	8	1.000	18.826	4	.000	131.128
69	4	4	.518	8	1.000	7.178	3	.000	32.181
70	1	3**	.015	8	.933	18.960	4	.042	25.185
71	4	4	.001	8	1.000	25.777	3	.000	63.468
72	9	5**	.000	8	.946	64.748	9	.054	70.473
73	3	3	.178	8	1.000	11.439	9	.000	31.814
74	4	4	.949	8	1.000	2.754	3	.000	33.145
75	3	3	.917	8	1.000	3.257	9	.000	38.867
76	6	6	.403	8	1.000	8.322	4	.000	47.869
77	9	9	.071	8	1.000	14.425	3	.000	30.587
78	2	2	.473	8	1.000	7.605	4	.000	70.217
79	4	4	.933	8	1.000	3.020	3	.000	27.336
80	3	3	.001	8	1.000	26.074	9	.000	49.019
81	3	3	.737	8	.999	5.194	9	.001	18.338
82	3	3	.002	8	.985	23.939	4	.015	32.373
83	2	2	.611	8	1.000	6.322	9	.000	55.484
84	8	8	.749	8	1.000	5.080	1	.000	166.666
85	5	5	.984	8	1.000	1.907	3	.000	128.024
86	4	4	.381	8	1.000	8.555	3	.000	42.407
87	1	1	.426	8	1.000	8.077	3	.000	75.281
88	5	5	.777	8	1.000	4.819	3	.000	157.569
89	4	4	.714	8	1.000	5.404	3	.000	26.041
90	4	4	.973	8	1.000	2.245	3	.000	33.468
91	4	4	.819	8	1.000	4.399	3	.000	34.601
92	5	5	.997	8	1.000	1.134	3	.000	133.133
93	4	4	.940	8	1.000	2.906	3	.000	33.884
94	3	3	.945	8	1.000	2.822	4	.000	37.510
95	2	2	.785	8	1.000	4.741	9	.000	53.876
96	1	3**	.043	8	.971	15.973	1	.029	23.022
97	6	6	.566	8	1.000	6.727	3	.000	34.968
98	9	9	.754	8	1.000	5.036	3	.000	49.709
99	4	4	.459	8	1.000	7.744	3	.000	30.290
100	6	6	.767	8	1.000	4.914	4	.000	24.200

101	6	6	.720	8	1.000	5.347	3	.000	38.309
102	6	6	.834	8	1.000	4.254	3	.000	28.486
103	6	6	.487	8	1.000	7.468	3	.000	63.813
104	3	3	.719	8	1.000	5.353	9	.000	23.669
105	9	9	.426	8	1.000	8.078	4	.000	48.380
106	6	6	.847	8	1.000	4.115	4	.000	29.515
107	4	4	.723	8	1.000	5.318	3	.000	34.774
108	2	2	.814	8	1.000	4.455	3	.000	60.794
109	7	7	.209	8	1.000	10.872	6	.000	43.070
110	4	4	.729	8	1.000	5.261	3	.000	29.472
111	8	8	.661	8	1.000	5.880	1	.000	95.083
112	1	1	.559	8	1.000	6.790	3	.000	34.515
113	8	8	.855	8	1.000	4.022	1	.000	115.483
114	8	8	.143	8	1.000	12.195	1	.000	81.390
115	9	9	.377	8	1.000	8.605	3	.000	29.094
116	3	3	.871	8	1.000	3.841	4	.000	23.371
117	3	3	.517	8	1.000	7.180	9	.000	23.899
118	9	9	.844	8	.999	4.147	3	.001	17.697
119	9	9	.718	8	1.000	5.367	3	.000	48.979
120	2	2	.764	8	1.000	4.935	9	.000	54.585
121	7	7	.989	8	1.000	1.692	3	.000	64.373
122	5	5	.612	8	1.000	6.318	7	.000	147.751
123	3	3	.706	8	1.000	5.469	4	.000	32.055
124	4	4	.969	8	1.000	2.328	6	.000	29.106
125	3	3	.672	8	1.000	5.775	9	.000	41.168
126	1	1	.206	8	1.000	10.925	4	.000	108.278
127	4	4	.615	8	1.000	6.285	3	.000	21.892
128	4	4	.565	8	.994	6.739	3	.005	17.160
129	9	9	.993	8	1.000	1.480	3	.000	29.960
130	3	3	.519	8	1.000	7.160	9	.000	36.156
131	7	7	.991	8	1.000	1.574	4	.000	71.272
132	7	7	.838	8	1.000	4.209	6	.000	49.520
133	7	7	.856	8	1.000	4.011	6	.000	68.272
134	7	7	.855	8	1.000	4.018	4	.000	45.167
135	7	7	.991	8	1.000	1.574	4	.000	71.272
136	4	4	.043	8	1.000	15.934	3	.000	71.908
137	2	2	.783	8	1.000	4.756	3	.000	52.639
138	4	4	.486	8	.990	7.478	3	.010	16.640
139	2	2	.733	8	1.000	5.230	3	.000	62.043
140	8	8	.182	8	1.000	11.357	1	.000	101.532
141	3	3	.828	8	1.000	4.314	9	.000	24.097
142	5	5	.176	8	1.000	11.474	3	.000	146.486
143	2	2	.916	8	1.000	3.276	3	.000	71.280
144	9	9	.120	8	1.000	12.773	3	.000	72.527
145	8	8	.804	8	1.000	4.552	1	.000	175.804
146	8	8	.950	8	1.000	2.733	1	.000	115.055
147	8	8	.997	8	1.000	1.108	1	.000	114.414
148	8	8	.702	8	1.000	5.510	1	.000	121.611
149	8	8	.997	8	1.000	1.108	1	.000	114.414
150	1	1	.385	8	1.000	8.516	4	.000	107.129

151	8	8	.997	8	1.000	1.108	1	.000	114.414
152	8	8	.999	8	1.000	.755	1	.000	144.293
153	8	8	.884	8	1.000	3.692	1	.000	165.916
154	4	4	.625	8	1.000	6.195	3	.000	40.957
155	4	4	.039	8	1.000	16.222	7	.000	32.821
156	1	1	.619	8	1.000	6.252	3	.000	91.297
157	5	5	.509	8	1.000	7.263	3	.000	157.707
158	5	5	.722	8	1.000	5.330	9	.000	131.342
159	3	3	.968	8	1.000	2.361	4	.000	32.953
160	7	7	.257	8	1.000	10.112	6	.000	53.298
161	9	9	.015	8	.837	19.065	2	.163	22.339
162	7	7	.958	8	1.000	2.568	3	.000	47.957
163	7	7	.997	8	1.000	1.204	3	.000	67.100
164	7	7	.997	8	1.000	1.204	3	.000	67.100
165	4	4	.644	8	1.000	6.028	3	.000	52.896
166	4	4	.191	8	1.000	11.196	6	.000	26.523
167	9	9	.864	8	1.000	3.928	3	.000	28.018
168	4	4	.804	8	1.000	4.552	6	.000	50.531
169	9	9	.247	8	.963	10.269	3	.037	16.789
170	3	3	.752	8	1.000	5.048	9	.000	45.661
171	3	3	.976	8	1.000	2.154	9	.000	33.080
172	3	3	.814	8	1.000	4.458	4	.000	31.912
173	2	2	.397	8	1.000	8.377	4	.000	60.164
174	4	4	.180	8	1.000	11.396	3	.000	55.001
175	1	1	.344	8	1.000	8.981	4	.000	107.850
176	7	7	.847	8	1.000	4.111	4	.000	74.377
177	7	7	.730	8	1.000	5.251	3	.000	89.393
178	8	8	.009	8	1.000	20.298	1	.000	142.251
179	7	7	.927	8	1.000	3.114	3	.000	41.049
180	7	7	.699	8	1.000	5.538	3	.000	88.419
181	8	8	.992	8	1.000	1.522	1	.000	139.271
182	8	8	.014	8	1.000	19.116	1	.000	101.326
183	6	6	.777	8	1.000	4.814	3	.000	26.056
184	6	6	.777	8	1.000	4.814	3	.000	26.056
185	6	6	.362	8	.980	8.766	4	.020	16.565
186	1	1	.667	8	1.000	5.827	3	.000	74.680
187	1	1	.001	8	1.000	26.830	3	.000	54.021
188	3	3	.870	8	1.000	3.852	4	.000	22.832
189	3	3	.940	8	1.000	2.913	9	.000	34.697
190	2	2	.090	8	1.000	13.694	6	.000	65.561
191	9	2**	.025	8	.722	17.541	9	.277	19.454
192	1	1	.139	8	1.000	12.282	7	.000	37.950
193	5	5	.267	8	1.000	9.971	7	.000	154.555
194	5	5	.765	8	1.000	4.930	4	.000	121.284
195	1	1	.588	8	1.000	6.531	9	.000	60.664
196	3	3	.873	8	1.000	3.824	4	.000	24.605
197	3	3	.220	8	.999	10.687	6	.001	23.743
198	6	6	.438	8	1.000	7.949	4	.000	42.451
199	6	6	.685	8	1.000	5.660	3	.000	29.705
200	2	2	.176	8	1.000	11.469	3	.000	62.779

201	5	5	.761	8	1.000	4.971	6	.000	152.089
202	9	5**	.000	8	.954	41.399	9	.046	47.463
203	5	5	.796	8	1.000	4.630	3	.000	132.866
204	5	5	.628	8	1.000	6.170	3	.000	130.648
205	3	3	.919	8	1.000	3.231	9	.000	21.408
206	4	4	.169	8	1.000	11.625	6	.000	38.442
207	4	4	.505	8	.989	7.294	3	.011	16.342
208	2	2	.628	8	1.000	6.172	3	.000	52.610
209	3	3	.059	8	1.000	15.016	4	.000	30.558
210	3	3	.988	8	1.000	1.718	9	.000	38.535
211	3	3	.791	8	1.000	4.677	9	.000	21.294
212	3	4**	.300	8	.530	9.530	3	.470	9.767
213	3	3	.981	8	1.000	1.991	4	.000	23.939
214	3	3	.963	8	1.000	2.476	4	.000	26.763
215	3	3	.457	8	1.000	7.764	9	.000	26.533
216	1	1	.536	8	1.000	7.003	4	.000	33.778
217	1	1	.258	8	1.000	10.094	3	.000	43.115
218	3	3	.861	8	1.000	3.956	9	.000	31.409
219	3	3	.992	8	1.000	1.526	9	.000	36.413
220	3	4**	.177	8	.976	11.467	3	.024	18.900
221	4	4	.723	8	1.000	5.319	6	.000	46.674
222	4	4	.996	8	1.000	1.275	3	.000	34.772
223	4	4	.978	8	1.000	2.083	3	.000	33.557
224	3	3	.984	8	1.000	1.885	9	.000	42.611
225	3	3	.734	8	1.000	5.215	4	.000	21.830
226	4	4	.896	8	1.000	3.542	6	.000	31.649
227	4	3**	.442	8	.977	7.917	4	.023	15.418
228	4	4	.480	8	1.000	7.535	3	.000	36.736
229	3	3	.491	8	1.000	7.435	6	.000	24.299
230	1	1	.146	8	1.000	12.123	2	.000	79.112
231	2	2	.000	8	1.000	31.756	6	.000	160.073
232	3	3	.760	8	1.000	4.974	4	.000	44.025
233	2	2	.732	8	1.000	5.235	9	.000	92.665
234	4	4	.558	8	1.000	6.807	6	.000	40.112
235	4	4	.387	8	.961	8.487	6	.038	14.931
236	9	9	.027	8	1.000	17.363	6	.000	64.537
237	2	2	.610	8	1.000	6.331	3	.000	40.292
238	2	2	.130	8	.995	12.495	3	.005	23.087
239	2	2	.999	8	1.000	.882	3	.000	59.083
240	2	2	1.000	8	1.000	.522	3	.000	55.275
241	3	3	.636	8	1.000	6.104	4	.000	46.041
242	2	2	.132	8	.999	12.459	4	.001	26.026
243	1	1	.763	8	1.000	4.954	3	.000	88.637
244	2	2	.820	8	1.000	4.393	3	.000	36.846
245	2	2	.939	8	1.000	2.923	3	.000	50.123
246	5	5	.972	8	1.000	2.261	3	.000	172.075
247	7	7	.811	8	1.000	4.488	3	.000	49.665
248	9	9	.673	8	1.000	5.768	3	.000	32.798
249	3	3	.615	8	.996	6.290	9	.004	17.225
250	9	9	.352	8	1.000	8.885	3	.000	24.133

251	9	9	.014	8	.918	19.147	3	.082	23.979
252	5	5	.557	8	1.000	6.814	3	.000	143.918
253	8	8	.914	8	1.000	3.303	1	.000	143.337
254	7	7	.203	8	1.000	10.972	3	.000	88.313
255	7	7	.965	8	1.000	2.438	3	.000	83.996
256	7	7	.965	8	1.000	2.438	3	.000	83.996
257	7	7	.965	8	1.000	2.438	3	.000	83.996
258	7	7	.780	8	1.000	4.785	6	.000	90.649
259	7	7	.861	8	1.000	3.954	6	.000	85.066
260	7	7	.965	8	1.000	2.438	3	.000	83.996
261	3	3	.256	8	.562	10.127	9	.438	10.624
262	7	7	.961	8	1.000	2.517	3	.000	71.266
263	7	7	.301	8	1.000	9.507	3	.000	44.837
264	5	5	.963	8	1.000	2.469	3	.000	172.331
265	2	2	.951	8	1.000	2.718	9	.000	78.500
266	4	4	.494	8	1.000	7.403	6	.000	23.561
267	9	9	.982	8	1.000	1.951	3	.000	37.417
268	9	9	.528	8	1.000	7.077	3	.000	54.619
269	9	9	.996	8	1.000	1.231	3	.000	37.811
270	3	3	.974	8	1.000	2.198	9	.000	30.822
271	9	9	.900	8	1.000	3.495	3	.000	45.232
272	7	7	.955	8	1.000	2.634	6	.000	48.463
273	2	2	.904	8	1.000	3.431	3	.000	51.974
274	9	9	.089	8	1.000	13.727	2	.000	34.976
275	2	2	.924	8	1.000	3.157	4	.000	62.554
276	1	1	.720	8	1.000	5.343	4	.000	90.516
277	2	2	.610	8	1.000	6.335	9	.000	87.138

Table 46: DFA 8 Cluster Results Tests of Equality of Group Means for the HCA Furthest Neighbour Jaccard Coefficient Model

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.591	26.636	7	269	.000
Customer	.267	105.624	7	269	.000
TargetSpecific	.946	2.206	7	269	.034
Unassociated	.225	132.483	7	269	.000
Received	.523	35.028	7	269	.000
Introduced	.782	10.686	7	269	.000
Sought	.666	19.267	7	269	.000
WebsiteorOnlineAuction	.811	8.978	7	269	.000
Face2Face	.842	7.186	7	269	.000
Text	.820	8.441	7	269	.000
Phone	.868	5.844	7	269	.000
Seminar	.909	3.847	7	269	.001
InternetForum	.837	7.475	7	269	.000
InternetPopUp	.791	10.169	7	269	.000
Email	.719	15.013	7	269	.000
Post	.753	12.581	7	269	.000
Advertisement	.742	13.376	7	269	.000
Fax	.955	1.829	7	269	.082
PrizeorMoney	.637	21.888	7	269	.000
HumanInteraction	.966	1.367	7	269	.219
FinancialReturn	.617	23.814	7	269	.000
Membership	.858	6.363	7	269	.000
AdviceorAssistance	.880	5.219	7	269	.000
Overpayment	.591	26.636	7	269	.000
Treatment	.890	4.752	7	269	.000
Employment	.156	208.016	7	269	.000
OpportunityForSelfOrOthers	.793	10.046	7	269	.000
Holiday	.930	2.879	7	269	.006
FinancialServices	.946	2.192	7	269	.035
GoodLuck	.973	1.064	7	269	.387
Property	.941	2.396	7	269	.022
Services	.922	3.239	7	269	.003
Merchandise	.659	19.898	7	269	.000
PartialPayment	.923	3.191	7	269	.003
Insight	.958	1.693	7	269	.111
Legal	.932	2.819	7	269	.008
FromFinancialInstitution	.872	5.626	7	269	.000
DetailUpdateorConfirmationRequired	.709	15.803	7	269	.000
GovernmentApproved	.941	2.399	7	269	.021
LoveAffectionConnection	.959	1.647	7	269	.122
GovernmentAgency	.971	1.137	7	269	.340
LargeReturn	.603	25.314	7	269	.000
Effective	.871	5.672	7	269	.000
RefundAvailable	.984	.630	7	269	.731
FraudulentActivity	.882	5.132	7	269	.000
ShareTips	.921	3.309	7	269	.002
NoCreditCheckRequired	.985	.601	7	269	.755
LittleorNoRisk	.853	6.602	7	269	.000
FromCorporateOrGovOfficial	.956	1.749	7	269	.098
QuickResponse	.947	2.134	7	269	.040
Confidentiality	.910	3.788	7	269	.001
PayupFrontCosts	.755	12.498	7	269	.000
ReceiveAndSendFunds	.778	10.983	7	269	.000
CallaPremiumNumber	.740	13.488	7	269	.000
TransferExcess	.633	22.321	7	269	.000
CompleteSaleoutsideofAuction	.923	3.191	7	269	.003
SendOntoOthers	.960	1.597	7	269	.136
RecruitOthers	.735	13.875	7	269	.000
SupplyPersonalInformation	.764	11.891	7	269	.000
SupplyBankAccDetails	.797	9.801	7	269	.000
Invest	.682	17.933	7	269	.000
MakeADonation	.912	3.720	7	269	.001
AlternativeShipment	.767	11.700	7	269	.000
Syntactic	.658	19.999	7	269	.000
Semantic	.680	18.120	7	269	.000
CompromisedWebsiteorFalseWebsite	.790	10.185	7	269	.000
DisguisedasInvoice	.946	2.194	7	269	.035
InferiorMerchandise	.881	5.195	7	269	.000
UseofFalsifiedForms	.852	6.661	7	269	.000
UseofParaphernalia	.827	8.038	7	269	.000
GoodsNeverSent	.803	9.415	7	269	.000
StoryBased	.806	9.242	7	269	.000
VerifiableStreetAddress	.984	.627	7	269	.733
LooksGenuine	.849	6.829	7	269	.000
ExploitLegitBusiness	.915	3.562	7	269	.001
Testimonials	.903	4.109	7	269	.000
RewardGreaterThanUpfrontCosts	.940	2.460	7	269	.018
FurtherContactbyEmailorPhone	.974	1.040	7	269	.403
PoliteBrokenEnglish	.983	.671	7	269	.696
FinancialGain	.326	79.384	7	269	.000
Information	.496	39.027	7	269	.000
Participation	.653	20.440	7	269	.000

Table 47: DFA 8 Cluster Results Variable Failing Tolerance Testing for the HCA Furthest Neighbour Jaccard Coefficient Model

Variables Failing Tolerance Test[a]

	Within-Groups Variance	Tolerance	Minimum Tolerance
Overpayment	.019	.000	.000

All variables passing the tolerance criteria are entered simultaneously.

a. Minimum tolerance level is .001.

Table 48: DFA 8 Cluster Results Eigenvalues for the HCA Furthest Neighbour Jaccard Coefficient Model

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	14.933[a]	36.5	36.5	.968
2	10.024[a]	24.5	61.0	.954
3	6.287[a]	15.4	76.4	.929
4	3.680[a]	9.0	85.4	.887
5	2.405[a]	5.9	91.2	.840
6	1.805[a]	4.4	95.7	.802
7	1.778[a]	4.3	100.0	.800

Table 49: DFA 8 Cluster Results Function Significance Tests for the HCA Furthest Neighbour Jaccard Coefficient Model

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 7	.000	2772.485	567	.000
2 through 7	.000	2131.598	480	.000
3 through 7	.001	1575.978	395	.000
4 through 7	.008	1116.197	312	.000
5 through 7	.038	758.926	231	.000
6 through 7	.128	475.314	152	.000
7	.360	236.541	75	.000

Table 50: DFA 8 Cluster Results Predicted Groups Memberships for the HCA Furthest Neighbour Jaccard Coefficient Model

		Highest Group					Second Highest Group		
			P(D>d | G=g)						
Case Number	Actual Group	Predicted Group	p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid
1	1	1	.223	7	1.000	9.429	8	.000	58.903
2	2	2	.980	7	1.000	1.574	3	.000	37.941
3	2	2	.587	7	1.000	5.603	3	.000	42.323
4	3	3	.275	7	1.000	8.700	2	.000	36.205
5	3	3	.002	7	1.000	23.312	1	.000	63.402
6	2	2	.286	7	1.000	8.562	5	.000	37.501
7	3	3	.423	7	1.000	7.061	8	.000	38.369
8	4	4	.058	7	1.000	13.654	8	.000	113.206
9	4	4	.871	7	1.000	3.145	2	.000	133.279
10	4	4	.935	7	1.000	2.387	3	.000	161.043
11	3	3	.155	7	1.000	10.648	5	.000	47.148
12	3	3	.919	7	1.000	2.609	8	.000	26.539
13	3	3	.956	7	1.000	2.074	2	.000	35.827
14	5	5	.058	7	1.000	13.624	2	.000	77.029
15	5	5	.120	7	1.000	11.443	3	.000	86.485
16	5	5	.752	7	1.000	4.238	3	.000	52.669
17	6	6	.981	7	1.000	1.534	2	.000	64.378
18	5	5	.128	7	1.000	11.244	6	.000	47.061
19	6	6	.056	7	.982	13.736	2	.009	23.048
20	2	2	.485	7	1.000	6.484	3	.000	35.889
21	3	3	.896	7	1.000	2.882	2	.000	29.174
22	7	7	.695	7	1.000	4.713	1	.000	149.643
23	8	8	.195	7	1.000	9.896	3	.000	64.581
24	7	7	.014	7	1.000	17.602	1	.000	148.120
25	6	6	.000	7	.864	28.625	1	.136	32.326
26	6	6	.712	7	1.000	4.575	5	.000	85.458
27	6	6	.508	7	1.000	6.271	2	.000	84.527
28	1	1	.762	7	1.000	4.155	3	.000	72.169
29	4	4	.780	7	1.000	3.997	6	.000	143.081
30	2	2	.410	7	1.000	7.184	3	.000	34.807
31	2	2	.913	7	1.000	2.679	3	.000	40.423
32	5	5	.290	7	.999	8.510	2	.001	21.590
33	8	8	.521	7	.998	6.163	3	.002	18.245
34	8	8	.944	7	1.000	2.255	3	.000	43.744
35	8	8	.949	7	1.000	2.182	3	.000	19.969
36	2	2	.459	7	1.000	6.718	8	.000	45.026
37	2	2	.938	7	1.000	2.354	5	.000	30.663
38	2	2	.682	7	1.000	4.819	3	.000	32.758
39	8	8	.204	7	1.000	9.729	3	.000	56.234
40	3	3	.955	7	1.000	2.092	8	.000	24.909
41	7	7	.928	7	1.000	2.485	1	.000	129.737
42	7	7	.279	7	1.000	8.653	1	.000	196.483
43	7	7	.927	7	1.000	2.502	1	.000	158.061
44	1	1	.383	7	1.000	7.452	2	.000	94.436
45	6	6	.023	7	.992	16.245	2	.006	26.449
46	6	6	.001	7	.541	24.485	2	.459	24.817
47	7	7	.767	7	1.000	4.112	1	.000	132.333
48	4	4	.992	7	1.000	1.129	3	.000	171.369
49	7	7	.999	7	1.000	.662	1	.000	132.933
50	6	6	.000	7	.777	50.357	4	.223	52.852

51	1	1	.195	7	1.000	9.893	6	.000	69.628
52	8	8	.405	7	1.000	7.230	2	.000	36.116
53	2	2	.991	7	1.000	1.191	5	.000	31.816
54	2	2	.483	7	.715	6.500	3	.285	8.343
55	6	6	.652	7	1.000	5.061	3	.000	63.864
56	8	8	.453	7	.834	6.775	3	.166	10.005
57	4	4	.423	7	1.000	7.058	6	.000	150.797
58	6	6	.993	7	1.000	1.115	3	.000	67.300
59	4	4	.517	7	1.000	6.196	5	.000	159.823
60	2	2	.374	7	.976	7.550	3	.024	15.000
61	2	2	.927	7	1.000	2.501	3	.000	26.185
62	3	3	.573	7	.985	5.721	8	.015	14.047
63	6	6	.631	7	1.000	5.237	2	.000	68.321
64	3	3	.119	7	.807	11.470	8	.193	14.334
65	2	2	.997	7	1.000	.879	3	.000	22.032
66	3	3	.828	7	1.000	3.564	2	.000	19.329
67	6	6	.121	7	1.000	11.437	3	.000	54.385
68	2	2	.110	7	1.000	11.715	3	.000	67.485
69	2	2	.623	7	1.000	5.303	3	.000	30.909
70	1	3**	.010	7	.898	18.484	2	.079	23.347
71	2	2	.001	7	1.000	24.940	3	.000	58.036
72	8	4**	.000	7	.954	64.636	8	.046	70.707
73	3	3	.131	7	1.000	11.190	8	.000	31.838
74	2	2	.882	7	1.000	3.027	3	.000	26.094
75	3	3	.935	7	1.000	2.391	2	.000	38.369
76	5	5	.554	7	1.000	5.879	2	.000	49.497
77	8	8	.057	7	1.000	13.672	3	.000	30.618
78	2	2	.280	7	1.000	8.635	3	.000	51.463
79	2	2	.836	7	1.000	3.493	3	.000	21.210
80	3	3	.000	7	1.000	26.095	8	.000	48.840
81	3	3	.658	7	.999	5.017	8	.001	18.283
82	3	3	.001	7	.781	23.757	2	.219	26.305
83	2	2	.426	7	1.000	7.026	3	.000	31.399
84	7	7	.660	7	1.000	4.999	1	.000	164.781
85	4	4	.970	7	1.000	1.811	2	.000	125.614
86	2	2	.292	7	1.000	8.488	3	.000	37.994
87	1	1	.377	7	1.000	7.514	3	.000	75.073
88	4	4	.831	7	1.000	3.541	3	.000	156.582
89	2	2	.594	7	.999	5.547	3	.001	20.140
90	2	2	.974	7	1.000	1.707	3	.000	23.211
91	2	2	.686	7	1.000	4.782	3	.000	27.009
92	4	4	.998	7	1.000	.779	2	.000	132.270
93	2	2	.885	7	1.000	3.000	3	.000	24.986
94	3	3	.918	7	1.000	2.619	2	.000	34.712
95	2	2	.664	7	1.000	4.963	3	.000	28.715
96	1	3**	.038	7	.961	14.858	1	.031	21.754
97	5	5	.606	7	1.000	5.441	3	.000	34.575
98	8	8	.662	7	1.000	4.985	3	.000	48.784
99	2	2	.391	7	.998	7.371	3	.002	19.525
100	5	5	.674	7	1.000	4.885	2	.000	20.977

101	5	5	.671	7	1.000	4.908	3	.000	38.320
102	5	5	.875	7	1.000	3.107	3	.000	28.142
103	5	5	.609	7	1.000	5.423	3	.000	62.883
104	3	3	.635	7	1.000	5.202	8	.000	23.571
105	8	8	.502	7	1.000	6.331	2	.000	44.389
106	5	5	.904	7	1.000	2.786	2	.000	30.197
107	2	2	.694	7	1.000	4.719	3	.000	23.493
108	2	2	.708	7	1.000	4.609	3	.000	31.899
109	6	6	.176	7	1.000	10.236	5	.000	41.979
110	2	2	.607	7	1.000	5.439	3	.000	24.090
111	7	7	.593	7	1.000	5.548	1	.000	92.365
112	1	1	.546	7	1.000	5.951	3	.000	33.912
113	7	7	.781	7	1.000	3.989	1	.000	114.814
114	7	7	.096	7	1.000	12.144	1	.000	79.737
115	8	8	.933	7	1.000	2.421	3	.000	25.934
116	3	3	.814	7	1.000	3.698	2	.000	19.841
117	3	3	.420	7	1.000	7.089	8	.000	23.720
118	8	8	.761	7	.999	4.158	3	.001	17.171
119	8	8	.630	7	1.000	5.247	3	.000	47.837
120	2	2	.676	7	1.000	4.865	8	.000	25.390
121	6	6	.990	7	1.000	1.244	3	.000	63.998
122	4	4	.530	7	1.000	6.087	6	.000	148.235
123	3	3	.651	7	1.000	5.070	2	.000	29.732
124	2	2	.951	7	1.000	2.158	5	.000	26.588
125	3	3	.587	7	1.000	5.602	2	.000	35.927
126	1	1	.251	7	1.000	9.029	2	.000	95.238
127	2	2	.566	7	.999	5.778	3	.001	19.294
128	2	2	.661	7	.996	4.990	3	.004	16.184
129	8	8	.983	7	1.000	1.485	3	.000	29.460
130	3	3	.524	7	1.000	6.134	2	.000	33.549
131	6	6	.992	7	1.000	1.169	2	.000	69.683
132	6	6	.754	7	1.000	4.222	5	.000	49.599
133	6	6	.798	7	1.000	3.840	5	.000	68.518
134	6	6	.781	7	1.000	3.987	2	.000	39.759
135	6	6	.992	7	1.000	1.169	2	.000	69.683
136	2	2	.047	7	1.000	14.228	3	.000	62.844
137	2	2	.795	7	1.000	3.863	3	.000	27.086
138	2	2	.388	7	.936	7.402	3	.064	12.782
139	2	2	.605	7	1.000	5.448	3	.000	26.454
140	7	7	.304	7	1.000	8.336	1	.000	101.114
141	3	3	.836	7	1.000	3.495	8	.000	24.092
142	4	4	.141	7	1.000	10.940	3	.000	146.628
143	2	2	.920	7	1.000	2.587	3	.000	29.841
144	8	8	.084	7	1.000	12.544	3	.000	71.080
145	7	7	.714	7	1.000	4.554	1	.000	174.842
146	7	7	.938	7	1.000	2.351	1	.000	112.061
147	7	7	.995	7	1.000	.991	1	.000	112.292
148	7	7	.601	7	1.000	5.485	1	.000	119.870
149	7	7	.995	7	1.000	.991	1	.000	112.292
150	1	1	.289	7	1.000	8.519	2	.000	103.610
151	7	7	.995	7	1.000	.991	1	.000	112.292

152	7	7	.998	7	1.000	.752	1	.000	143.068
153	7	7	.853	7	1.000	3.325	1	.000	165.817
154	2	2	.797	7	1.000	3.852	3	.000	25.439
155	2	2	.024	7	.989	16.137	6	.011	25.166
156	1	1	.543	7	1.000	5.974	3	.000	91.241
157	4	4	.410	7	1.000	7.185	3	.000	158.101
158	4	4	.715	7	1.000	4.545	8	.000	129.644
159	3	3	.984	7	1.000	1.452	2	.000	31.770
160	6	6	.185	7	1.000	10.061	5	.000	53.375
161	8	8	.175	7	.989	10.250	2	.011	19.280
162	6	6	.923	7	1.000	2.557	3	.000	48.065
163	6	6	.993	7	1.000	1.115	3	.000	67.300
164	6	6	.993	7	1.000	1.115	3	.000	67.300
165	2	2	.780	7	1.000	4.002	8	.000	35.921
166	2	2	.137	7	.967	11.045	5	.033	17.818
167	8	8	.850	7	1.000	3.361	3	.000	25.855
168	2	2	.870	7	1.000	3.155	5	.000	39.795
169	8	8	.193	7	.966	9.924	3	.031	16.803
170	3	3	.672	7	1.000	4.902	2	.000	43.706
171	3	3	.961	7	1.000	1.981	2	.000	32.469
172	3	3	.728	7	1.000	4.437	2	.000	26.041
173	2	2	.282	7	1.000	8.608	3	.000	58.186
174	2	2	.164	7	1.000	10.453	3	.000	51.728
175	1	1	.271	7	1.000	8.755	2	.000	105.936
176	6	6	.769	7	1.000	4.091	2	.000	70.720
177	6	6	.689	7	1.000	4.759	2	.000	87.031
178	7	7	.005	7	1.000	20.249	1	.000	141.770
179	6	6	.877	7	1.000	3.083	2	.000	37.585
180	6	6	.603	7	1.000	5.469	3	.000	88.575
181	7	7	.991	7	1.000	1.195	1	.000	136.579
182	7	7	.008	7	1.000	19.089	1	.000	100.772
183	5	5	.737	7	1.000	4.367	3	.000	26.032
184	5	5	.737	7	1.000	4.367	3	.000	26.032
185	5	5	.488	7	.996	6.449	2	.004	17.314
186	1	1	.650	7	1.000	5.083	8	.000	73.717
187	1	1	.000	7	1.000	26.142	3	.000	53.293
188	3	3	.804	7	.999	3.789	2	.001	18.972
189	3	3	.910	7	1.000	2.721	8	.000	33.300
190	2	2	.067	7	.999	13.224	5	.001	26.595
191	8	8	.179	7	.829	10.175	2	.154	13.539
192	1	1	.091	7	1.000	12.312	6	.000	38.078
193	4	4	.218	7	1.000	9.513	6	.000	154.243
194	4	4	.760	7	1.000	4.173	2	.000	118.976
195	1	1	.483	7	1.000	6.502	8	.000	60.456
196	3	3	.860	7	1.000	3.258	2	.000	22.779
197	3	3	.151	7	.998	10.717	5	.002	23.636
198	5	5	.433	7	1.000	6.964	2	.000	42.650
199	5	5	.632	7	1.000	5.232	3	.000	29.663
200	2	2	.117	7	1.000	11.538	3	.000	33.694

201	4	4	.830	7	1.000	3.546	2	.000	142.503
202	8	4**	.000	7	.950	41.499	8	.050	47.393
203	4	4	.732	7	1.000	4.405	3	.000	133.230
204	4	4	.523	7	1.000	6.140	3	.000	131.126
205	3	3	.940	7	1.000	2.324	8	.000	21.316
206	2	2	.165	7	1.000	10.449	5	.000	28.101
207	2	2	.585	7	.992	5.619	3	.008	15.239
208	2	2	.632	7	1.000	5.228	3	.000	25.474
209	3	3	.160	7	1.000	10.547	2	.000	31.954
210	3	3	.987	7	1.000	1.346	2	.000	34.756
211	3	3	.720	7	1.000	4.504	8	.000	21.228
212	3	2**	.348	7	.586	7.831	3	.414	8.525
213	3	3	.971	7	1.000	1.790	2	.000	20.618
214	3	3	.947	7	1.000	2.208	2	.000	24.059
215	3	3	.599	7	1.000	5.504	8	.000	21.488
216	1	1	.458	7	1.000	6.726	2	.000	31.013
217	1	1	.183	7	1.000	10.100	2	.000	42.632
218	3	3	.908	7	1.000	2.735	2	.000	28.309
219	3	3	.988	7	1.000	1.320	2	.000	34.593
220	3	3	.146	7	.682	10.841	2	.318	12.366
221	2	2	.905	7	1.000	2.770	5	.000	32.979
222	2	2	.992	7	1.000	1.133	3	.000	30.085
223	2	2	.956	7	1.000	2.076	3	.000	28.470
224	3	3	.966	7	1.000	1.886	2	.000	39.431
225	3	3	.692	7	.999	4.736	2	.001	19.615
226	2	2	.873	7	1.000	3.125	5	.000	29.388
227	2	3**	.341	7	.806	7.908	2	.194	10.752
228	2	2	.850	7	1.000	3.361	3	.000	36.549
229	3	3	.427	7	.999	7.022	2	.001	21.798
230	1	1	.266	7	1.000	8.813	2	.000	63.653
231	2	2	.208	7	1.000	9.673	5	.000	44.142
232	3	3	.715	7	1.000	4.549	2	.000	42.222
233	2	2	.963	7	1.000	1.936	3	.000	38.401
234	2	2	.803	7	1.000	3.799	5	.000	39.479
235	2	2	.251	7	.876	9.023	5	.120	12.998
236	8	8	.033	7	1.000	15.210	5	.000	57.735
237	2	2	.907	7	1.000	2.756	3	.000	26.467
238	2	2	.904	7	.999	2.783	3	.001	16.390
239	2	2	.993	7	1.000	1.080	3	.000	27.085
240	2	2	1.000	7	1.000	.428	3	.000	26.083
241	3	3	.607	7	1.000	5.433	2	.000	44.882
242	2	2	.966	7	1.000	1.876	3	.000	26.683
243	1	1	.665	7	1.000	4.962	3	.000	88.948
244	2	2	.969	7	1.000	1.826	3	.000	18.651
245	2	2	.932	7	1.000	2.429	3	.000	21.908
246	4	4	.959	7	1.000	2.007	2	.000	168.402
247	6	6	.758	7	1.000	4.186	3	.000	49.389
248	8	8	.570	7	1.000	5.744	2	.000	31.692
249	3	3	.506	7	.994	6.292	8	.006	16.639
250	8	8	.266	7	.999	8.814	3	.001	23.953

251	8	8	.011	7	.945	18.229	3	.055	23.905
252	4	4	.509	7	1.000	6.269	3	.000	143.679
253	7	7	.892	7	1.000	2.918	1	.000	143.362
254	6	6	.139	7	1.000	10.996	3	.000	88.621
255	6	6	.934	7	1.000	2.401	2	.000	82.175
256	6	6	.934	7	1.000	2.401	2	.000	82.175
257	6	6	.934	7	1.000	2.401	2	.000	82.175
258	6	6	.692	7	1.000	4.740	5	.000	90.970
259	6	6	.786	7	1.000	3.943	5	.000	85.171
260	6	6	.934	7	1.000	2.401	2	.000	82.175
261	3	8**	.202	7	.544	9.770	3	.456	10.119
262	6	6	.946	7	1.000	2.236	2	.000	70.097
263	6	6	.672	7	1.000	4.905	2	.000	40.824
264	4	4	.950	7	1.000	2.159	2	.000	171.672
265	2	2	.957	7	1.000	2.052	3	.000	39.184
266	2	2	.337	7	.984	7.955	5	.016	16.196
267	8	8	.963	7	1.000	1.941	3	.000	36.751
268	8	8	.420	7	1.000	7.082	3	.000	53.941
269	8	8	.993	7	1.000	1.113	3	.000	37.618
270	3	3	.969	7	1.000	1.823	2	.000	25.587
271	8	8	.857	7	1.000	3.286	3	.000	45.218
272	6	6	.929	7	1.000	2.469	5	.000	48.085
273	2	2	.877	7	1.000	3.081	3	.000	30.521
274	8	8	.481	7	1.000	6.514	2	.000	30.917
275	2	2	.836	7	1.000	3.488	3	.000	36.423
276	1	1	.715	7	1.000	4.544	2	.000	80.928
277	2	2	.657	7	1.000	5.028	3	.000	42.184

Table 51: DFA 7 Cluster Results Tests of Equality of Group Means for the HCA Furthest Neighbour Jaccard Coefficient Model

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.904	4.758	6	270	.000
Customer	.208	123.176	6	270	.000
TargetSpecific	.946	2.552	6	270	.020
Unassociated	.411	64.366	6	270	.000
Received	.559	35.474	6	270	.000
Introduced	.786	12.242	6	270	.000
Sought	.667	22.466	6	270	.000
WebsiteorOnlineAuction	.824	9.589	6	270	.000
Face2Face	.853	7.736	6	270	.000
Text	.820	9.885	6	270	.000
Phone	.935	3.124	6	270	.006
Seminar	.910	4.465	6	270	.000
InternetForum	.837	8.753	6	270	.000
InternetPopUp	.791	11.908	6	270	.000
Email	.773	13.185	6	270	.000
Post	.807	10.787	6	270	.000
Advertisement	.753	14.752	6	270	.000
Fax	.955	2.142	6	270	.049
PrizeorMoney	.638	25.566	6	270	.000
HumanInteraction	.966	1.601	6	270	.147
FinancialReturn	.626	26.842	6	270	.000
Membership	.860	7.307	6	270	.000
AdviceorAssistance	.909	4.504	6	270	.000
Overpayment	.904	4.758	6	270	.000
Treatment	.895	5.300	6	270	.000
Employment	.156	243.508	6	270	.000
OpportunityForSelfOrOthers	.793	11.764	6	270	.000
Holiday	.932	3.259	6	270	.004
FinancialServices	.946	2.550	6	270	.020
GoodLuck	.973	1.234	6	270	.289
Property	.942	2.747	6	270	.013
Services	.927	3.536	6	270	.002
Merchandise	.672	21.982	6	270	.000
PartialPayment	.923	3.736	6	270	.001
Insight	.958	1.959	6	270	.072
Legal	.932	3.301	6	270	.004
FromFinancialInstitution	.874	6.511	6	270	.000
DetailUpdateorConfirmationRequired	.709	18.475	6	270	.000
GovernmentApproved	.951	2.332	6	270	.033
LoveAffectionConnection	.959	1.928	6	270	.076
GovernmentAgency	.981	.883	6	270	.508
LargeReturn	.617	27.938	6	270	.000
Effective	.886	5.807	6	270	.000
RefundAvailable	.988	.557	6	270	.764
FraudulentActivity	.882	6.009	6	270	.000
ShareTips	.932	3.250	6	270	.004
NoCreditCheckRequired	.986	.649	6	270	.691
LittleorNoRisk	.878	6.253	6	270	.000
FromCorporateOrGovOfficial	.956	2.048	6	270	.060
QuickResponse	.949	2.426	6	270	.027
Confidentiality	.910	4.436	6	270	.000
PayupFrontCosts	.758	14.388	6	270	.000
ReceiveAndSendFunds	.778	12.822	6	270	.000
CallaPremiumNumber	.740	15.795	6	270	.000
TransferExcess	.916	4.125	6	270	.001
CompleteSaleoutsideofAuction	.923	3.736	6	270	.001
SendOntoOthers	.960	1.870	6	270	.086
RecruitOthers	.735	16.187	6	270	.000
SupplyPersonalInformation	.764	13.921	6	270	.000
SupplyBankAccDetails	.797	11.468	6	270	.000
Invest	.688	20.437	6	270	.000
MakeADonation	.912	4.336	6	270	.000
AlternativeShipment	.767	13.700	6	270	.000
Syntactic	.659	23.246	6	270	.000
Semantic	.680	21.219	6	270	.000
CompromisedWebsiteorFalseWebsite	.791	11.905	6	270	.000
DisguisedasInvoice	.965	1.650	6	270	.134
InferiorMerchandise	.919	3.975	6	270	.001
UseofFalsifiedForms	.853	7.777	6	270	.000
UseofParaphernalia	.854	7.708	6	270	.000
GoodsNeverSent	.804	10.967	6	270	.000
StoryBased	.809	10.598	6	270	.000
VerifiableStreetAddress	.990	.469	6	270	.831
LooksGenuine	.880	6.125	6	270	.000
ExploitLegitBusiness	.924	3.727	6	270	.001
Testimonials	.921	3.870	6	270	.001
RewardGreaterThanUpfrontCosts	.945	2.631	6	270	.017
FurtherContactbyEmailorPhone	.974	1.218	6	270	.297
PoliteBrokenEnglish	.984	.732	6	270	.624
FinancialGain	.326	92.958	6	270	.000
Information	.498	45.412	6	270	.000
Participation	.653	23.893	6	270	.000

Table 52: DFA 7 Cluster Results Variable Failing Tolerance Testing for the HCA Furthest Neighbour Jaccard Coefficient Model

Variables Failing Tolerance Test[a]

	Within-Groups Variance	Tolerance	Minimum Tolerance
Overpayment	.029	.000	.000

Table 53: DFA 7 Cluster Results Eigenvalues for the HCA Furthest Neighbour Jaccard Coefficient Model

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	10.838[a]	31.6	31.6	.957
2	9.662[a]	28.2	59.8	.952
3	6.284[a]	18.3	78.2	.929
4	3.677[a]	10.7	88.9	.887
5	1.997[a]	5.8	94.7	.816
6	1.805[a]	5.3	100.0	.802

Table 54: DFA 7 Cluster Results Function Significance Tests for the HCA Furthest Neighbour Jaccard Coefficient Model

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 6	.000	2434.828	486	.000
2 through 6	.000	1861.482	400	.000
3 through 6	.003	1312.420	316	.000
4 through 6	.025	851.730	234	.000
5 through 6	.119	493.856	154	.000
6	.357	239.246	76	.000

Table 55: DFA 7 Cluster Results Predicted Groups Memberships for the HCA Furthest Neighbour Jaccard Coefficient Model

		Highest Group					Second Highest Group		
			P(D>d | G=g)						
Case Number	Actual Group	Predicted Group	p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid
1	1	1	.266	6	.871	7.636	7	.124	11.541
2	2	2	.961	6	1.000	1.478	1	.000	33.822
3	2	2	.469	6	1.000	5.601	1	.000	38.544
4	1	1	.686	6	1.000	3.935	2	.000	35.923
5	1	1	.004	6	1.000	19.263	2	.000	67.347
6	2	2	.214	6	1.000	8.337	4	.000	37.374
7	1	1	.534	6	1.000	5.081	7	.000	30.006
8	3	3	.058	6	1.000	12.173	7	.000	111.122
9	3	3	.795	6	1.000	3.113	2	.000	133.748
10	3	3	.902	6	1.000	2.180	1	.000	159.874
11	1	1	.155	6	1.000	9.354	4	.000	40.644
12	1	1	.826	6	1.000	2.861	7	.000	22.785
13	1	1	.924	6	1.000	1.951	2	.000	30.009
14	4	4	.034	6	1.000	13.624	2	.000	77.263
15	4	4	.087	6	1.000	11.045	1	.000	85.370
16	4	4	.652	6	1.000	4.185	1	.000	50.267
17	5	5	.960	6	1.000	1.496	2	.000	64.264
18	4	4	.195	6	1.000	8.642	5	.000	46.227
19	5	5	.033	6	.799	13.750	1	.192	16.607
20	2	2	.379	6	1.000	6.405	1	.000	31.477
21	1	1	.854	6	1.000	2.625	2	.000	21.125
22	6	6	.734	6	1.000	3.577	1	.000	157.560
23	7	7	.145	6	1.000	9.554	1	.000	64.656
24	6	6	.080	6	1.000	11.285	5	.000	127.022
25	5	5	.030	6	1.000	13.950	1	.000	34.356
26	5	5	.632	6	1.000	4.328	1	.000	85.397
27	5	5	.629	6	1.000	4.352	2	.000	80.535
28	1	1	.660	6	1.000	4.125	2	.000	39.029
29	3	3	.679	6	1.000	3.980	5	.000	142.810
30	2	2	.487	6	1.000	5.453	1	.000	35.179
31	2	2	.849	6	1.000	2.668	1	.000	37.705
32	4	4	.212	6	.999	8.371	2	.001	21.510
33	7	7	.428	6	.989	5.960	1	.011	15.048
34	7	7	.899	6	1.000	2.216	1	.000	41.652
35	7	7	.978	6	1.000	1.172	1	.000	20.557
36	2	2	.856	6	1.000	2.614	1	.000	32.201
37	2	2	.887	6	1.000	2.331	1	.000	27.021
38	2	2	.625	6	1.000	4.384	1	.000	31.946
39	7	7	.412	6	1.000	6.096	1	.000	57.360
40	1	1	.969	6	1.000	1.340	7	.000	23.795
41	6	6	.950	6	1.000	1.632	1	.000	137.866
42	6	6	.219	6	1.000	8.275	2	.000	202.115
43	6	6	.896	6	1.000	2.244	1	.000	167.307
44	1	1	.882	6	1.000	2.381	2	.000	34.647
45	5	5	.018	6	.749	15.292	1	.237	17.593
46	5	2**	.003	6	.758	19.897	5	.241	22.192
47	6	6	.662	6	1.000	4.112	2	.000	139.056
48	3	3	.980	6	1.000	1.133	1	.000	168.721
49	6	6	.996	6	1.000	.605	1	.000	142.258
50	5	5	.000	6	.613	50.057	3	.387	50.976

51	1	1	.047	6	1.000	12.771	5	.000	32.715
52	7	7	.347	6	1.000	6.729	1	.000	32.039
53	2	2	.991	6	1.000	.821	1	.000	26.089
54	2	2	.473	6	.763	5.569	1	.237	7.906
55	5	5	.977	6	1.000	1.197	1	.000	63.536
56	7	7	.396	6	.822	6.252	1	.178	9.308
57	3	3	.487	6	1.000	5.453	5	.000	150.862
58	5	5	.991	6	1.000	.847	1	.000	63.027
59	3	3	.648	6	1.000	4.210	1	.000	154.109
60	2	2	.278	6	.792	7.491	1	.208	10.166
61	2	2	.950	6	1.000	1.630	1	.000	24.754
62	1	1	.404	6	.919	6.177	7	.080	11.050
63	5	5	.531	6	1.000	5.100	1	.000	62.933
64	1	7**	.086	6	.601	11.083	1	.398	11.908
65	2	2	.995	6	1.000	.653	1	.000	20.339
66	1	1	.721	6	.999	3.673	2	.001	17.245
67	5	5	.087	6	1.000	11.060	1	.000	49.861
68	2	2	.112	6	1.000	10.308	1	.000	58.007
69	2	2	.664	6	1.000	4.095	1	.000	22.583
70	1	1	.367	6	.999	6.521	2	.001	21.121
71	2	2	.001	6	1.000	23.458	1	.000	49.815
72	7	3**	.000	6	.965	64.223	7	.035	70.828
73	1	1	.489	6	.993	5.441	7	.007	15.427
74	2	2	.876	6	1.000	2.432	1	.000	25.167
75	1	1	.895	6	1.000	2.253	2	.000	30.977
76	4	4	.478	6	1.000	5.525	2	.000	49.319
77	7	7	.051	6	1.000	12.514	1	.000	31.671
78	2	2	.237	6	1.000	8.021	1	.000	50.094
79	2	2	.759	6	1.000	3.384	1	.000	19.157
80	1	1	.002	6	1.000	21.254	7	.000	39.996
81	1	1	.550	6	.986	4.952	7	.014	13.432
82	1	1	.002	6	.537	21.044	2	.463	21.342
83	2	2	.327	6	1.000	6.929	1	.000	27.379
84	6	6	.645	6	1.000	4.235	1	.000	173.351
85	3	3	.940	6	1.000	1.770	2	.000	126.011
86	2	2	.250	6	1.000	7.846	1	.000	30.718
87	1	1	.607	6	1.000	4.518	2	.000	34.458
88	3	3	.878	6	1.000	2.416	1	.000	156.320
89	2	2	.599	6	.999	4.579	1	.001	18.062
90	2	2	.956	6	1.000	1.555	1	.000	21.429
91	2	2	.602	6	1.000	4.556	1	.000	21.255
92	3	3	.995	6	1.000	.698	1	.000	131.483
93	2	2	.883	6	1.000	2.371	1	.000	24.188
94	1	1	.825	6	1.000	2.873	2	.000	32.271
95	2	2	.561	6	1.000	4.869	1	.000	24.356
96	1	1	.890	6	1.000	2.302	2	.000	21.585
97	4	4	.613	6	1.000	4.469	1	.000	33.938
98	7	7	.584	6	1.000	4.692	1	.000	45.521
99	2	2	.367	6	.932	6.524	1	.068	11.769
100	4	4	.634	6	1.000	4.316	2	.000	20.482

101	4	4	.714	6	1.000	3.725	1	.000	30.073
102	4	4	.831	6	1.000	2.821	1	.000	26.050
103	4	4	.563	6	1.000	4.853	1	.000	56.836
104	1	1	.633	6	.999	4.322	7	.001	19.334
105	7	7	.449	6	1.000	5.772	1	.000	44.164
106	4	4	.918	6	1.000	2.015	2	.000	29.544
107	2	2	.583	6	.999	4.698	1	.001	19.297
108	2	2	.844	6	1.000	2.715	1	.000	31.905
109	5	5	.120	6	1.000	10.122	4	.000	41.391
110	2	2	.665	6	1.000	4.087	1	.000	24.058
111	6	6	.574	6	1.000	4.765	1	.000	100.327
112	1	1	.928	6	1.000	1.907	2	.000	25.740
113	6	6	.841	6	1.000	2.738	1	.000	123.613
114	6	6	.072	6	1.000	11.590	1	.000	87.937
115	7	7	.948	6	1.000	1.665	1	.000	26.041
116	1	1	.616	6	.996	4.447	2	.004	15.681
117	1	1	.334	6	.999	6.859	2	.000	23.271
118	7	7	.660	6	.997	4.126	1	.003	16.031
119	7	7	.828	6	1.000	2.844	1	.000	38.950
120	2	2	.571	6	1.000	4.788	7	.000	25.016
121	5	5	.981	6	1.000	1.123	1	.000	57.983
122	3	3	.476	6	1.000	5.546	5	.000	148.785
123	1	1	.798	6	1.000	3.089	2	.000	29.304
124	2	2	.918	6	1.000	2.022	4	.000	26.537
125	1	1	.400	6	1.000	6.210	2	.000	34.285
126	1	1	.929	6	1.000	1.891	2	.000	26.322
127	2	2	.480	6	.980	5.509	1	.020	13.305
128	2	2	.575	6	.992	4.756	1	.008	14.364
129	7	7	.961	6	1.000	1.474	1	.000	28.361
130	1	1	.413	6	1.000	6.093	2	.000	28.986
131	5	5	.981	6	1.000	1.123	1	.000	64.833
132	5	5	.651	6	1.000	4.193	1	.000	48.064
133	5	5	.887	6	1.000	2.328	1	.000	56.809
134	5	5	.684	6	1.000	3.947	2	.000	39.119
135	5	5	.981	6	1.000	1.123	1	.000	64.833
136	2	2	.057	6	1.000	12.252	1	.000	63.600
137	2	2	.717	6	1.000	3.702	1	.000	25.131
138	2	2	.371	6	.947	6.486	1	.053	12.244
139	2	2	.669	6	1.000	4.057	1	.000	26.788
140	6	6	.250	6	1.000	7.840	2	.000	94.192
141	1	1	.909	6	.999	2.109	7	.001	15.987
142	3	3	.189	6	1.000	8.735	1	.000	147.272
143	2	2	.873	6	1.000	2.459	1	.000	25.010
144	7	7	.054	6	1.000	12.359	1	.000	71.156
145	6	6	.633	6	1.000	4.324	2	.000	180.243
146	6	6	.936	6	1.000	1.808	1	.000	120.472
147	6	6	.988	6	1.000	.940	2	.000	121.353
148	6	6	.607	6	1.000	4.520	1	.000	127.743
149	6	6	.988	6	1.000	.940	2	.000	121.353
150	1	1	.987	6	1.000	.963	2	.000	28.929

151	6	6	.988	6	1.000	.940	2	.000	121.353
152	6	6	.994	6	1.000	.725	2	.000	151.242
153	6	6	.807	6	1.000	3.016	2	.000	174.548
154	2	2	.857	6	1.000	2.598	1	.000	25.538
155	2	2	.013	6	.986	16.189	5	.013	24.776
156	1	1	.809	6	1.000	2.999	7	.000	31.628
157	3	3	.317	6	1.000	7.046	1	.000	155.600
158	3	3	.613	6	1.000	4.470	1	.000	129.775
159	1	1	.931	6	1.000	1.871	2	.000	26.666
160	5	5	.128	6	1.000	9.925	4	.000	53.444
161	7	7	.114	6	.988	10.269	2	.012	19.067
162	5	5	.868	6	1.000	2.508	1	.000	41.104
163	5	5	.991	6	1.000	.847	1	.000	63.027
164	5	5	.991	6	1.000	.847	1	.000	63.027
165	2	2	.740	6	1.000	3.530	1	.000	31.872
166	2	2	.086	6	.968	11.066	4	.032	17.861
167	7	7	.762	6	1.000	3.367	1	.000	24.232
168	2	2	.798	6	1.000	3.084	1	.000	35.592
169	7	7	.175	6	.720	8.969	1	.278	10.872
170	1	1	.955	6	1.000	1.560	2	.000	28.945
171	1	1	.953	6	1.000	1.597	2	.000	26.419
172	1	1	.702	6	1.000	3.815	2	.000	19.570
173	2	2	.210	6	1.000	8.397	1	.000	56.869
174	2	2	.112	6	1.000	10.311	1	.000	48.936
175	1	1	.948	6	1.000	1.668	2	.000	36.245
176	5	5	.677	6	1.000	3.998	1	.000	68.063
177	5	5	.579	6	1.000	4.732	1	.000	83.368
178	6	6	.026	6	1.000	14.369	5	.000	128.872
179	5	5	.802	6	1.000	3.051	1	.000	35.196
180	5	5	.518	6	1.000	5.201	1	.000	83.872
181	6	6	.995	6	1.000	.687	1	.000	145.225
182	6	6	.005	6	1.000	18.420	4	.000	92.262
183	4	4	.747	6	1.000	3.478	1	.000	25.166
184	4	4	.747	6	1.000	3.478	1	.000	25.166
185	4	4	.412	6	.996	6.102	2	.004	16.994
186	1	1	.310	6	1.000	7.115	7	.000	30.015
187	1	1	.001	6	1.000	23.915	7	.000	52.118
188	1	1	.828	6	.999	2.846	2	.001	17.835
189	1	1	.923	6	1.000	1.961	7	.000	33.052
190	2	2	.069	6	.999	11.711	4	.001	25.100
191	7	7	.118	6	.795	10.173	2	.179	13.156
192	1	1	.109	6	.995	10.382	2	.004	21.452
193	3	3	.166	6	1.000	9.145	5	.000	153.159
194	3	3	.654	6	1.000	4.169	2	.000	119.399
195	1	1	.167	6	.998	9.111	7	.002	21.777
196	1	1	.707	6	1.000	3.775	2	.000	21.085
197	1	1	.097	6	.996	10.718	4	.003	22.219
198	4	4	.502	6	1.000	5.328	2	.000	41.119
199	4	4	.521	6	1.000	5.182	1	.000	25.672
200	2	2	.143	6	1.000	9.586	1	.000	30.963

201	3	3	.832	6	1.000	2.809	2	.000	142.356
202	7	3**	.000	6	.950	41.537	7	.050	47.446
203	3	3	.646	6	1.000	4.228	1	.000	131.774
204	3	3	.482	6	1.000	5.495	1	.000	130.915
205	1	1	.891	6	1.000	2.288	7	.000	20.219
206	2	2	.110	6	1.000	10.368	4	.000	28.094
207	2	2	.495	6	.875	5.387	1	.125	9.283
208	2	2	.597	6	1.000	4.593	1	.000	24.650
209	1	1	.085	6	.999	11.103	2	.001	24.601
210	1	1	.989	6	1.000	.901	2	.000	26.897
211	1	1	.882	6	1.000	2.373	7	.000	21.270
212	1	1	.319	6	.534	7.025	2	.466	7.295
213	1	1	.894	6	1.000	2.260	2	.000	17.486
214	1	1	.823	6	1.000	2.885	2	.000	19.497
215	1	1	.453	6	.999	5.740	7	.001	20.192
216	1	1	.660	6	.993	4.122	2	.007	14.006
217	1	1	.169	6	1.000	9.077	2	.000	28.387
218	1	1	.905	6	1.000	2.149	2	.000	27.248
219	1	1	.959	6	1.000	1.507	2	.000	30.287
220	1	2**	.166	6	.784	9.144	1	.216	11.727
221	2	2	.908	6	1.000	2.126	4	.000	32.431
222	2	2	.994	6	1.000	.713	1	.000	23.961
223	2	2	.985	6	1.000	1.018	1	.000	20.057
224	1	1	.982	6	1.000	1.083	2	.000	32.179
225	1	1	.554	6	.998	4.923	2	.002	17.651
226	2	2	.811	6	1.000	2.985	4	.000	29.335
227	2	1**	.580	6	.953	4.721	2	.047	10.761
228	2	2	.816	6	1.000	2.945	1	.000	30.743
229	1	1	.362	6	.997	6.573	2	.002	18.976
230	1	1	.126	6	.944	9.970	2	.056	15.631
231	2	2	.191	6	1.000	8.705	4	.000	43.278
232	1	1	.557	6	1.000	4.899	2	.000	39.893
233	2	2	.939	6	1.000	1.775	1	.000	33.314
234	2	2	.740	6	1.000	3.526	4	.000	39.352
235	2	2	.172	6	.850	9.026	4	.116	13.011
236	7	7	.019	6	1.000	15.172	4	.000	57.879
237	2	2	.870	6	1.000	2.483	1	.000	21.050
238	2	2	.836	6	.994	2.781	1	.006	12.901
239	2	2	.985	6	1.000	1.010	1	.000	22.417
240	2	2	.999	6	1.000	.302	1	.000	24.020
241	1	1	.623	6	1.000	4.400	2	.000	43.714
242	2	2	.984	6	1.000	1.037	1	.000	18.837
243	1	1	.811	6	1.000	2.984	7	.000	33.098
244	2	2	.947	6	.999	1.673	1	.001	16.621
245	2	2	.900	6	1.000	2.208	1	.000	20.414
246	3	3	.919	6	1.000	2.011	2	.000	169.017
247	5	5	.749	6	1.000	3.462	1	.000	46.149
248	7	7	.918	6	1.000	2.014	1	.000	21.713
249	1	1	.468	6	.995	5.613	7	.005	16.337
250	7	7	.247	6	.994	7.877	1	.006	18.195

251	7	7	.023	6	.975	14.664	1	.025	22.004
252	3	3	.392	6	1.000	6.284	1	.000	140.683
253	6	6	.865	6	1.000	2.535	1	.000	152.192
254	5	5	.128	6	1.000	9.920	1	.000	84.493
255	5	5	.884	6	1.000	2.360	1	.000	78.367
256	5	5	.884	6	1.000	2.360	1	.000	78.367
257	5	5	.884	6	1.000	2.360	1	.000	78.367
258	5	5	.578	6	1.000	4.738	1	.000	85.445
259	5	5	.729	6	1.000	3.614	1	.000	83.301
260	5	5	.884	6	1.000	2.360	1	.000	78.367
261	1	7**	.206	6	.742	8.460	1	.258	10.571
262	5	5	.932	6	1.000	1.865	1	.000	60.873
263	5	5	.647	6	1.000	4.218	1	.000	37.627
264	3	3	.911	6	1.000	2.095	1	.000	170.967
265	2	2	.937	6	1.000	1.808	1	.000	34.302
266	2	2	.244	6	.983	7.917	4	.016	16.195
267	7	7	.938	6	1.000	1.792	1	.000	36.017
268	7	7	.382	6	1.000	6.381	1	.000	54.852
269	7	7	.989	6	1.000	.912	1	.000	37.490
270	1	1	.959	6	1.000	1.507	2	.000	18.483
271	7	7	.954	6	1.000	1.584	1	.000	37.853
272	5	5	.873	6	1.000	2.458	1	.000	46.701
273	2	2	.881	6	1.000	2.385	1	.000	29.300
274	7	7	.371	6	1.000	6.489	2	.000	30.535
275	2	2	.797	6	1.000	3.096	1	.000	35.094
276	1	1	.960	6	1.000	1.500	2	.000	23.830
277	2	2	.905	6	1.000	2.156	1	.000	28.693

Table 56: DFA 6 Cluster Results Tests of Equality of Group Means for the HCA Furthest Neighbour Jaccard Coefficient Model

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.904	5.730	5	271	.000
Customer	.268	147.983	5	271	.000
TargetSpecific	.948	2.946	5	271	.013
Unassociated	.411	77.526	5	271	.000
Received	.560	42.599	5	271	.000
Introduced	.900	6.010	5	271	.000
Sought	.724	20.651	5	271	.000
WebsiteorOnlineAuction	.824	11.548	5	271	.000
Face2Face	.911	5.323	5	271	.000
Text	.823	11.675	5	271	.000
Phone	.936	3.678	5	271	.003
Seminar	.955	2.547	5	271	.028
InternetForum	.840	10.360	5	271	.000
InternetPopUp	.791	14.343	5	271	.000
Email	.774	15.871	5	271	.000
Post	.807	12.955	5	271	.000
Advertisement	.841	10.282	5	271	.000
Fax	.955	2.580	5	271	.027
PrizeorMoney	.638	30.745	5	271	.000
HumanInteraction	.966	1.916	5	271	.092
FinancialReturn	.627	32.243	5	271	.000
Membership	.932	3.961	5	271	.002
AdviceorAssistance	.923	4.552	5	271	.001
Overpayment	.904	5.730	5	271	.000
Treatment	.895	6.383	5	271	.000
Employment	.592	37.307	5	271	.000
OpportunityForSelfOrOthers	.802	13.380	5	271	.000
Holiday	.932	3.925	5	271	.002
FinancialServices	.954	2.641	5	271	.024
GoodLuck	.973	1.487	5	271	.194
Property	.968	1.773	5	271	.119
Services	.947	3.051	5	271	.011
Merchandise	.673	26.360	5	271	.000
PartialPayment	.923	4.500	5	271	.001
Insight	.958	2.360	5	271	.041
Legal	.955	2.550	5	271	.028
FromFinancialInstitution	.874	7.819	5	271	.000
DetailUpdateorConfirmationRequired	.712	21.892	5	271	.000
GovernmentApproved	.974	1.440	5	271	.210
LoveAffectionConnection	.961	2.181	5	271	.057
GovernmentAgency	.982	.981	5	271	.430
LargeReturn	.786	14.765	5	271	.000
Effective	.894	6.418	5	271	.000
RefundAvailable	.990	.538	5	271	.747
FraudulentActivity	.884	7.094	5	271	.000
ShareTips	.962	2.160	5	271	.058
NoCreditCheckRequired	.989	.622	5	271	.683
LittleorNoRisk	.885	7.054	5	271	.000
FromCorporateOrGovOfficial	.958	2.380	5	271	.039
QuickResponse	.957	2.459	5	271	.034
Confidentiality	.912	5.256	5	271	.000
PayupFrontCosts	.766	16.576	5	271	.000
ReceiveAndSendFunds	.904	5.733	5	271	.000
CallaPremiumNumber	.740	19.024	5	271	.000
TransferExcess	.919	4.794	5	271	.000
CompleteSaleoutsideofAuction	.923	4.500	5	271	.001
SendOntoOthers	.960	2.252	5	271	.050
RecruitOthers	.791	14.344	5	271	.000
SupplyPersonalInformation	.770	16.165	5	271	.000
SupplyBankAccDetails	.822	11.733	5	271	.000
Invest	.829	11.165	5	271	.000
MakeADonation	.912	5.223	5	271	.000
AlternativeShipment	.767	16.501	5	271	.000
Syntactic	.660	27.912	5	271	.000
Semantic	.680	25.557	5	271	.000
CompromisedWebsiteorFalseWebsite	.792	14.276	5	271	.000
DisguisedasInvoice	.965	1.987	5	271	.081
InferiorMerchandise	.920	4.727	5	271	.000
UseofFalsifiedForms	.854	9.302	5	271	.000
UseofParaphernalia	.899	6.096	5	271	.000
GoodsNeverSent	.804	13.209	5	271	.000
StoryBased	.810	12.685	5	271	.000
VerifiableStreetAddress	.990	.565	5	271	.727
LooksGenuine	.906	5.625	5	271	.000
ExploitLegitBusiness	.926	4.356	5	271	.001
Testimonials	.937	3.662	5	271	.003
RewardGreaterThanUpfrontCosts	.953	2.678	5	271	.022
FurtherContactbyEmailorPhone	.974	1.455	5	271	.205
PoliteBrokenEnglish	.987	.721	5	271	.608
FinancialGain	.338	106.062	5	271	.000
Information	.511	51.815	5	271	.000
Participation	.706	22.540	5	271	.000

Table 57: DFA 6 Cluster Results Variable Failing Tolerance Testing for the HCA Furthest Neighbour Jaccard Coefficient Model

Variables Failing Tolerance Test[a]

	Within-Groups Variance	Tolerance	Minimum Tolerance
Overpayment	.029	.000	.000

Table 58: DFA 6 Cluster Results Eigenvalues for the HCA Furthest Neighbour Jaccard Coefficient Model

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	10.400[a]	41.2	41.2	.955
2	6.328[a]	25.1	66.3	.929
3	4.212[a]	16.7	82.9	.899
4	2.503[a]	9.9	92.9	.845
5	1.805[a]	7.1	100.0	.802

Table 59: DFA 6 Cluster Results Function Significance Tests for the HCA Furthest Neighbour Jaccard Coefficient Model

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 5	.000	1943.980	405	.000
2 through 5	.003	1378.174	320	.000
3 through 5	.020	915.092	237	.000
4 through 5	.102	531.256	156	.000
5	.357	239.788	77	.000

Table 60: DFA 6 Cluster Results Predicted Groups Memberships for the HCA Furthest Neighbour Jaccard Coefficient Model

		Highest Group					Second Highest Group		
			P(D>d | G=g)						
Case Number	Actual Group	Predicted Group	p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid
1	1	1	.279	5	.953	6.289	3	.040	12.634
2	2	2	.940	5	1.000	1.254	1	.000	33.863
3	2	2	.508	5	1.000	4.290	1	.000	37.851
4	1	1	.610	5	1.000	3.592	2	.000	35.848
5	1	1	.003	5	1.000	18.221	2	.000	66.714
6	2	2	.285	5	1.000	6.224	4	.000	34.221
7	1	1	.824	5	1.000	2.177	3	.000	33.125
8	3	3	.159	5	1.000	7.944	1	.000	55.325
9	3	3	.865	5	1.000	1.883	1	.000	24.206
10	3	3	.974	5	1.000	.840	1	.000	28.110
11	1	1	.390	5	1.000	5.216	4	.000	38.418
12	1	1	.915	5	1.000	1.485	3	.000	24.863
13	1	1	.930	5	1.000	1.350	2	.000	29.797
14	4	4	.018	5	1.000	13.652	2	.000	77.296
15	4	4	.052	5	1.000	10.992	1	.000	84.825
16	4	4	.527	5	1.000	4.159	1	.000	50.154
17	5	5	.965	5	1.000	.968	2	.000	60.578
18	4	4	.130	5	1.000	8.508	5	.000	46.002
19	5	5	.019	5	.723	13.552	1	.267	15.544
20	2	2	.270	5	1.000	6.389	1	.000	31.391
21	1	1	.812	5	1.000	2.258	2	.000	21.025
22	6	6	.827	5	1.000	2.160	1	.000	155.355
23	3	3	.233	5	1.000	6.838	1	.000	57.056
24	6	6	.074	5	1.000	10.050	5	.000	127.476
25	5	5	.020	5	1.000	13.367	1	.000	30.352
26	5	5	.706	5	1.000	2.962	1	.000	78.484
27	5	5	.531	5	1.000	4.127	2	.000	79.914
28	1	1	.566	5	1.000	3.882	2	.000	38.972
29	3	3	.446	5	1.000	4.755	1	.000	34.707
30	2	2	.563	5	1.000	3.904	1	.000	33.077
31	2	2	.790	5	1.000	2.409	1	.000	37.775
32	4	4	.150	5	.998	8.110	2	.002	20.858
33	3	3	.280	5	.969	6.277	1	.031	13.188
34	3	3	.950	5	1.000	1.144	1	.000	31.741
35	3	3	.850	5	.999	1.995	1	.001	15.626
36	2	2	.761	5	1.000	2.600	1	.000	32.176
37	2	2	.939	5	1.000	1.260	1	.000	26.482
38	2	2	.522	5	1.000	4.191	1	.000	31.550
39	3	3	.814	5	1.000	2.246	1	.000	41.535
40	1	1	.936	5	1.000	1.291	3	.000	22.985
41	6	6	.898	5	1.000	1.629	1	.000	138.201
42	6	6	.168	5	1.000	7.795	2	.000	202.595
43	6	6	.821	5	1.000	2.195	1	.000	167.421
44	1	1	.833	5	1.000	2.113	2	.000	34.630
45	5	5	.009	5	.550	15.247	1	.433	15.726
46	5	2**	.003	5	.885	18.133	5	.112	22.264
47	6	6	.534	5	1.000	4.107	5	.000	138.673
48	3	3	.954	5	1.000	1.098	1	.000	38.400
49	6	6	.993	5	1.000	.485	1	.000	142.074
50	5	5	.005	5	1.000	16.728	3	.000	41.114

51	1	1	.036	5	1.000	11.931	5	.000	32.295
52	3	3	.220	5	.999	7.006	1	.000	22.763
53	2	2	.978	5	1.000	.788	1	.000	26.123
54	2	2	.363	5	.774	5.452	1	.226	7.911
55	5	5	.951	5	1.000	1.132	1	.000	61.627
56	3	3	.243	5	.650	6.715	1	.350	7.953
57	3	3	.298	5	1.000	6.088	2	.000	38.677
58	5	5	.974	5	1.000	.844	1	.000	60.674
59	3	3	.433	5	1.000	4.860	1	.000	40.385
60	2	2	.341	5	.722	5.658	1	.278	7.568
61	2	2	.907	5	1.000	1.553	1	.000	24.822
62	1	1	.365	5	.938	5.434	3	.062	10.877
63	5	5	.428	5	1.000	4.901	1	.000	61.946
64	1	1	.280	5	.969	6.279	3	.030	13.213
65	2	2	.985	5	1.000	.652	1	.000	20.310
66	1	1	.595	5	.999	3.686	2	.001	17.240
67	5	5	.057	5	1.000	10.709	1	.000	49.207
68	2	2	.114	5	1.000	8.881	1	.000	56.148
69	2	2	.567	5	1.000	3.879	1	.000	22.196
70	1	1	.938	5	1.000	1.272	2	.000	16.989
71	2	2	.000	5	1.000	23.356	1	.000	49.959
72	3	3	.000	5	1.000	27.581	1	.000	102.153
73	1	1	.489	5	.998	4.432	3	.002	16.492
74	2	2	.871	5	1.000	1.839	1	.000	24.279
75	1	1	.815	5	1.000	2.239	2	.000	31.065
76	4	4	.410	5	1.000	5.051	2	.000	48.604
77	3	3	.022	5	.992	13.162	1	.008	22.746
78	2	2	.177	5	1.000	7.645	1	.000	50.109
79	2	2	.744	5	1.000	2.716	1	.000	18.143
80	1	1	.001	5	1.000	21.305	3	.000	37.931
81	1	1	.471	5	.986	4.566	3	.014	13.068
82	1	1	.001	5	.516	21.073	2	.484	21.199
83	2	2	.291	5	1.000	6.155	1	.000	27.026
84	6	6	.533	5	1.000	4.118	1	.000	173.957
85	3	3	.938	5	1.000	1.271	1	.000	38.795
86	2	2	.167	5	1.000	7.818	1	.000	30.774
87	1	1	.507	5	1.000	4.300	2	.000	34.429
88	3	3	.912	5	1.000	1.506	1	.000	30.093
89	2	2	.471	5	.999	4.570	1	.001	18.117
90	2	2	.933	5	1.000	1.320	1	.000	21.428
91	2	2	.790	5	1.000	2.410	1	.000	18.421
92	3	3	.995	5	1.000	.406	1	.000	31.192
93	2	2	.808	5	1.000	2.286	1	.000	24.085
94	1	1	.729	5	1.000	2.809	2	.000	32.159
95	2	2	.526	5	1.000	4.167	3	.000	23.137
96	1	1	.813	5	1.000	2.254	2	.000	21.649
97	4	4	.565	5	1.000	3.895	1	.000	33.766
98	3	3	.998	5	1.000	.258	1	.000	26.550
99	2	2	.416	5	.948	5.000	1	.052	10.818
100	4	4	.665	5	1.000	3.228	2	.000	20.028

101	4	4	.594	5	1.000	3.697	1	.000	29.644
102	4	4	.792	5	1.000	2.397	1	.000	24.573
103	4	4	.483	5	1.000	4.480	1	.000	56.969
104	1	1	.592	5	1.000	3.710	3	.000	20.342
105	3	3	.695	5	1.000	3.032	2	.000	27.174
106	4	4	.900	5	1.000	1.614	2	.000	29.553
107	2	2	.492	5	.999	4.409	1	.001	18.756
108	2	2	.751	5	1.000	2.666	1	.000	32.023
109	5	5	.078	5	1.000	9.912	4	.000	40.157
110	2	2	.709	5	1.000	2.942	1	.000	23.464
111	6	6	.529	5	1.000	4.144	1	.000	100.598
112	1	1	.865	5	1.000	1.886	2	.000	25.824
113	6	6	.749	5	1.000	2.683	1	.000	123.873
114	6	6	.044	5	1.000	11.411	1	.000	87.847
115	3	3	.837	5	1.000	2.083	1	.000	18.191
116	1	1	.631	5	.997	3.452	2	.003	15.189
117	1	1	.347	5	.993	5.603	3	.007	15.630
118	3	3	.418	5	.978	4.984	1	.022	12.538
119	3	3	.910	5	1.000	1.530	1	.000	27.215
120	2	2	.569	5	1.000	3.868	3	.000	24.560
121	5	5	.960	5	1.000	1.036	1	.000	55.027
122	3	3	.227	5	1.000	6.910	1	.000	42.724
123	1	1	.797	5	1.000	2.365	2	.000	28.980
124	2	2	.854	5	1.000	1.964	4	.000	26.357
125	1	1	.304	5	1.000	6.025	2	.000	34.353
126	1	1	.874	5	1.000	1.813	2	.000	26.308
127	2	2	.396	5	.974	5.163	1	.024	12.599
128	2	2	.449	5	.991	4.736	1	.008	14.258
129	3	3	.912	5	1.000	1.509	1	.000	19.229
130	1	1	.573	5	1.000	3.836	2	.000	25.981
131	5	5	.956	5	1.000	1.081	1	.000	63.203
132	5	5	.537	5	1.000	4.089	1	.000	46.100
133	5	5	.900	5	1.000	1.609	1	.000	56.524
134	5	5	.613	5	1.000	3.569	2	.000	38.849
135	5	5	.956	5	1.000	1.081	1	.000	63.203
136	2	2	.035	5	1.000	11.980	1	.000	63.676
137	2	2	.591	5	1.000	3.713	1	.000	25.163
138	2	2	.299	5	.952	6.077	1	.048	12.055
139	2	2	.639	5	1.000	3.400	1	.000	26.568
140	6	6	.164	5	1.000	7.855	2	.000	94.443
141	1	1	.886	5	.999	1.727	3	.001	16.055
142	3	3	.359	5	.994	5.491	1	.006	15.799
143	2	2	.935	5	1.000	1.295	1	.000	24.433
144	3	3	.747	5	1.000	2.697	1	.000	44.372
145	6	6	.524	5	1.000	4.175	2	.000	180.828
146	6	6	.884	5	1.000	1.741	1	.000	120.335
147	6	6	.969	5	1.000	.920	2	.000	121.765
148	6	6	.492	5	1.000	4.413	1	.000	127.512
149	6	6	.969	5	1.000	.920	2	.000	121.765
150	1	1	.966	5	1.000	.956	2	.000	28.911

151	6	6	.969	5	1.000	.920	2	.000	121.765
152	6	6	.984	5	1.000	.679	2	.000	151.789
153	6	6	.710	5	1.000	2.934	2	.000	174.905
154	2	2	.795	5	1.000	2.377	1	.000	25.511
155	2	2	.006	5	.979	16.164	5	.021	23.845
156	1	1	.730	5	1.000	2.807	3	.000	32.059
157	3	3	.163	5	1.000	7.883	1	.000	45.678
158	3	3	.560	5	1.000	3.924	1	.000	46.513
159	1	1	.896	5	1.000	1.641	2	.000	26.202
160	5	5	.078	5	1.000	9.899	4	.000	52.840
161	3	3	.039	5	.743	11.679	2	.257	13.806
162	5	5	.851	5	1.000	1.985	1	.000	40.579
163	5	5	.974	5	1.000	.844	1	.000	60.674
164	5	5	.974	5	1.000	.844	1	.000	60.674
165	2	2	.660	5	1.000	3.263	3	.000	29.452
166	2	2	.058	5	.972	10.670	4	.028	17.770
167	3	3	.622	5	.992	3.510	1	.008	13.102
168	2	2	.687	5	1.000	3.085	1	.000	35.596
169	3	3	.177	5	.781	7.651	1	.218	10.205
170	1	1	.930	5	1.000	1.348	2	.000	28.982
171	1	1	.904	5	1.000	1.577	3	.000	26.501
172	1	1	.618	5	1.000	3.539	2	.000	19.496
173	2	2	.138	5	1.000	8.356	1	.000	57.062
174	2	2	.163	5	1.000	7.878	1	.000	45.745
175	1	1	.896	5	1.000	1.641	3	.000	36.089
176	5	5	.952	5	1.000	1.124	1	.000	67.825
177	5	5	.464	5	1.000	4.622	1	.000	82.304
178	6	6	.019	5	1.000	13.540	5	.000	125.878
179	5	5	.707	5	1.000	2.957	1	.000	33.871
180	5	5	.473	5	1.000	4.554	1	.000	83.463
181	6	6	.988	5	1.000	.599	1	.000	145.721
182	6	6	.002	5	1.000	18.472	4	.000	92.594
183	4	4	.685	5	1.000	3.095	1	.000	23.787
184	4	4	.685	5	1.000	3.095	1	.000	23.787
185	4	4	.302	5	.995	6.043	2	.005	16.728
186	1	1	.250	5	1.000	6.631	3	.000	31.143
187	1	1	.000	5	1.000	22.567	3	.000	54.813
188	1	1	.947	5	1.000	1.183	2	.000	16.772
189	1	1	.864	5	1.000	1.891	3	.000	32.896
190	2	2	.045	5	.999	11.324	4	.001	25.060
191	3	2**	.063	5	.577	10.458	3	.361	11.392
192	1	1	.074	5	.993	10.039	2	.004	20.875
193	3	3	.080	5	1.000	9.837	1	.000	46.707
194	3	3	.733	5	1.000	2.785	1	.000	39.332
195	1	1	.110	5	.997	8.988	3	.003	20.401
196	1	1	.582	5	1.000	3.779	2	.000	21.134
197	1	1	.113	5	.991	8.905	4	.008	18.521
198	4	4	.407	5	1.000	5.072	2	.000	41.222
199	4	4	.405	5	1.000	5.089	1	.000	25.638
200	2	2	.103	5	1.000	9.150	1	.000	30.196

201	3	3	.747	5	1.000	2.692	2	.000	35.940
202	3	3	.415	5	1.000	5.007	1	.000	47.205
203	3	3	.533	5	1.000	4.113	1	.000	40.619
204	3	3	.414	5	1.000	5.019	1	.000	45.100
205	1	1	.853	5	.999	1.970	3	.001	15.925
206	2	2	.067	5	1.000	10.287	4	.000	27.734
207	2	2	.378	5	.863	5.325	1	.137	9.012
208	2	2	.720	5	1.000	2.869	1	.000	22.342
209	1	1	.050	5	.999	11.061	2	.001	24.551
210	1	1	.970	5	1.000	.897	2	.000	26.916
211	1	1	.825	5	1.000	2.173	3	.000	21.223
212	1	1	.221	5	.510	6.991	2	.490	7.068
213	1	1	.822	5	1.000	2.194	2	.000	17.531
214	1	1	.726	5	1.000	2.833	2	.000	19.539
215	1	1	.595	5	1.000	3.689	3	.000	22.381
216	1	1	.638	5	.994	3.401	2	.006	13.649
217	1	1	.146	5	1.000	8.199	2	.000	27.909
218	1	1	.846	5	1.000	2.020	2	.000	26.943
219	1	1	.919	5	1.000	1.451	2	.000	30.134
220	1	2**	.125	5	.751	8.634	1	.249	10.838
221	2	2	.835	5	1.000	2.103	4	.000	32.519
222	2	2	.983	5	1.000	.696	1	.000	23.999
223	2	2	.974	5	1.000	.853	1	.000	19.710
224	1	1	.958	5	1.000	1.055	2	.000	32.194
225	1	1	.620	5	.999	3.519	2	.001	16.792
226	2	2	.708	5	1.000	2.945	4	.000	29.111
227	2	1**	.455	5	.950	4.686	2	.050	10.558
228	2	2	.789	5	1.000	2.416	1	.000	29.948
229	1	1	.315	5	.998	5.915	2	.002	18.676
230	1	1	.096	5	.931	9.355	2	.069	14.549
231	2	2	.160	5	1.000	7.932	4	.000	43.046
232	1	1	.449	5	1.000	4.738	2	.000	39.904
233	2	2	.907	5	1.000	1.554	1	.000	32.895
234	2	2	.709	5	1.000	2.944	4	.000	38.292
235	2	2	.134	5	.861	8.425	4	.096	12.812
236	3	3	.048	5	1.000	11.196	4	.000	37.303
237	2	2	.862	5	1.000	1.905	1	.000	20.874
238	2	2	.738	5	.993	2.754	1	.007	12.757
239	2	2	.975	5	1.000	.834	1	.000	22.439
240	2	2	.998	5	1.000	.257	1	.000	23.883
241	1	1	.512	5	1.000	4.268	2	.000	43.471
242	2	2	.962	5	1.000	1.008	1	.000	18.744
243	1	1	.796	5	1.000	2.372	3	.000	34.587
244	2	2	.916	5	.999	1.478	1	.001	16.189
245	2	2	.915	5	1.000	1.480	1	.000	20.079
246	3	3	.966	5	1.000	.956	1	.000	33.706
247	5	5	.651	5	1.000	3.320	1	.000	44.601
248	3	3	.849	5	1.000	2.002	1	.000	19.020
249	1	1	.394	5	.996	5.185	3	.004	16.188
250	3	1**	.180	5	.649	7.600	3	.350	8.835

251	3	3	.006	5	.665	16.423	1	.335	17.794
252	3	3	.500	5	1.000	4.348	1	.000	22.547
253	6	6	.774	5	1.000	2.517	1	.000	152.308
254	5	5	.078	5	1.000	9.913	1	.000	82.844
255	5	5	.800	5	1.000	2.341	1	.000	76.341
256	5	5	.800	5	1.000	2.341	1	.000	76.341
257	5	5	.800	5	1.000	2.341	1	.000	76.341
258	5	5	.447	5	1.000	4.749	1	.000	83.167
259	5	5	.608	5	1.000	3.603	1	.000	81.423
260	5	5	.800	5	1.000	2.341	1	.000	76.341
261	1	3**	.163	5	.697	7.888	1	.303	9.554
262	5	5	.916	5	1.000	1.473	1	.000	59.758
263	5	5	.521	5	1.000	4.202	1	.000	35.888
264	3	3	.963	5	1.000	.994	1	.000	35.475
265	2	2	.939	5	1.000	1.263	1	.000	34.167
266	2	2	.162	5	.982	7.894	4	.017	16.033
267	3	3	.909	5	1.000	1.534	1	.000	30.181
268	3	3	.747	5	1.000	2.697	1	.000	38.974
269	3	3	.998	5	1.000	.291	1	.000	28.268
270	1	1	.917	5	1.000	1.462	2	.000	18.543
271	3	3	.955	5	1.000	1.084	1	.000	30.116
272	5	5	.786	5	1.000	2.436	1	.000	43.990
273	2	2	.805	5	1.000	2.306	1	.000	29.096
274	3	3	.193	5	1.000	7.387	2	.000	24.910
275	2	2	.703	5	1.000	2.983	1	.000	34.920
276	1	1	.947	5	1.000	1.173	2	.000	23.760
277	2	2	.842	5	1.000	2.049	1	.000	28.793

Table 61: DFA 9 Cluster Results Tests of Equality of Group Means for the HCA Within Groups Linkage Jaccard Coefficient Model

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.258	96.146	8	268	.000
Customer	.321	70.900	0	260	.000
TargetSpecific	.910	3.299	0	260	.001
Unassociated	.231	111.767	0	260	.000
Received	.515	31.542	0	260	.000
Introduced	.698	14.504	8	268	.000
Sought	.546	27.824	8	268	.000
WebsiteorOnlineAuction	.725	12.712	8	268	.000
Face2Face	.799	8.415	8	268	.000
Text	.952	1.696	8	268	.099
Phone	.729	12.440	8	268	.000
Seminar	.903	3.606	8	268	.001
InternetForum	.729	12.462	8	268	.000
InternetPopUp	.546	27.887	8	268	.000
Email	.727	12.586	8	268	.000
Post	.791	8.866	8	268	.000
Advertisement	.618	20.701	8	268	.000
Fax	.968	1.099	8	268	.364
PrizeorMoney	.735	12.077	8	268	.000
HumanInteraction	.951	1.741	8	268	.089
FinancialReturn	.623	20.304	8	268	.000
Membership	.906	3.479	0	260	.001
AdviceorAssistance	.909	3.364	0	260	.001
Overpayment	.250	96.146	0	260	.000
Treatment	.273	09.320	0	260	.000
Employment	.551	27.294	8	268	.000
OpportunityForSelfOrOthers	.878	4.672	8	268	.000
Holiday	.955	1.578	8	268	.131
FinancialServices	.974	.878	8	268	.536
GoodLuck	.962	1.306	8	268	.240
Property	.980	.670	8	268	.718
Services	.957	1.517	8	268	.151
Merchandise	.543	28.183	8	268	.000
PartialPayment	.930	2.530	8	268	.011
Insight	.954	1.630	8	268	.116
Legal	.946	1.924	8	268	.057
FromFinancialInstitution	.717	13.253	8	268	.000
DetailUpdateorConfirmationRequired	.687	15.282	8	268	.000
GovernmentApproved	.953	1.661	8	268	.108
LoveAffectionConnection	.976	.832	8	268	.575
GovernmentAgency	.950	1.746	8	268	.088
LargeReturn	.510	32.123	0	260	.000
Effective	.431	44.311	0	260	.000
RefundAvailable	.953	1.651	0	260	.111
FraudulentActivity	.604	15.400	0	260	.000
ShareTips	.024	7.166	0	260	.000
NoCreditCheckRequired	.966	1.170	8	268	.318
LittleorNoRisk	.798	8.505	8	268	.000
FromCorporateOrGovOfficial	.934	2.357	8	268	.018
QuickResponse	.934	2.370	8	268	.018
Confidentiality	.915	3.102	8	268	.002
PayupFrontCosts	.748	11.311	8	268	.000
ReceiveAndSendFunds	.826	7.055	8	268	.000
CallaPremiumNumber	.928	2.607	8	268	.009
TransferExcess	.333	67.046	8	268	.000
CompleteSaleoutsideofAuction	.930	2.530	8	268	.011
SendOntoOthers	.948	1.843	8	268	.069
RecruitOthers	.787	9.064	8	268	.000
SupplyPersonalInformation	.733	12.176	8	268	.000
SupplyBankAccDetails	.779	9.494	8	268	.000
Invest	.683	15.570	8	268	.000
MakeADonation	.917	3.013	8	268	.003
AlternativeShipment	.786	9.107	8	268	.000
Syntactic	.412	47.889	8	268	.000
Semantic	.335	66.513	8	268	.000
CompromisedWebsiteorFalseWebsite	.749	11.197	8	268	.000
DisguisedasInvoice	.949	1.798	8	268	.078
InferiorMerchandise	.837	6.523	8	268	.000
UseofFalsifiedForms	.616	20.897	8	268	.000
UseofParaphernalia	.827	7.010	8	268	.000
GoodsNeverSent	.812	7.770	8	268	.000
StoryBased	.060	5.001	0	260	.000
VerifiableStreetAddress	.905	.521	0	260	.040
LooksGenuine	.777	9.590	0	260	.000
ExploitLegitBusiness	.044	6.190	0	260	.000
Testimonials	.604	21.979	0	260	.000
RewardGreaterThanUpfrontCosts	.963	1.207	0	260	.250
FurtherContactbyEmailorPhone	.937	2.240	8	268	.025
PoliteBrokenEnglish	.964	1.238	8	268	.277
FinancialGain	.425	45.347	8	268	.000
Information	.524	30.476	8	268	.000
Participation	.779	9.524	8	268	.000

Table 62: DFA 9 Cluster Results Variable Failing Tolerance Testing for the HCA Within Groups Linkage Jaccard Coefficient Model

Variables Failing Tolerance Test[a]

	Within-Groups Variance	Tolerance	Minimum Tolerance
Overpayment	.008	.000	.000

Table 63: DFA 9 Cluster Results Eigenvalues for the HCA Within Groups Linkage Jaccard Coefficient Model

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	13.899[a]	31.5	31.5	.966
2	6.739[a]	15.2	46.7	.933
3	5.821[a]	13.2	59.9	.924
4	5.049[a]	11.4	71.3	.914
5	4.070[a]	9.2	80.5	.896
6	3.799[a]	8.6	89.1	.890
7	2.789[a]	6.3	95.4	.858
8	2.026[a]	4.6	100.0	.818

Table 64: DFA 9 Cluster Results Function Significance Tests for the HCA Within Groups Linkage Jaccard Coefficient Model

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 8	.000	3256.759	648	.000
2 through 8	.000	2632.761	560	.000
3 through 8	.000	2160.079	474	.000
4 through 8	.001	1716.565	390	.000
5 through 8	.004	1300.783	308	.000
6 through 8	.018	925.803	228	.000
7 through 8	.087	563.486	150	.000
8	.330	255.769	74	.000

Table 65: DFA 9 Cluster Results Predicted Groups Memberships for the HCA Within Groups Linkage Jaccard Coefficient Model

Case Number	Actual Group	Highest Group: Predicted Group	Highest Group: P(D>d | G=g) p	Highest Group: P(D>d | G=g) df	Highest Group: P(G=g | D=d)	Highest Group: Squared Mahalanobis Distance to Centroid	Second Highest Group: Group	Second Highest Group: P(G=g | D=d)	Second Highest Group: Squared Mahalanobis Distance to Centroid
1	1	1	.120	8	.979	12.780	2	.020	20.600
2	2	2	1.000	8	1.000	.644	3	.000	18.029
3	3	3	.545	8	1.000	6.926	2	.000	45.863
4	1	1	.903	8	1.000	3.449	2	.000	53.065
5	2	2	.004	8	1.000	22.827	3	.000	67.068
6	4	4	.125	8	1.000	12.640	3	.000	45.291
7	2	2	.267	8	1.000	9.979	3	.000	52.562
8	5	5	.026	8	1.000	17.415	1	.000	38.881
9	3	3	.083	8	.890	13.943	5	.110	18.121
10	5	5	.813	8	1.000	4.466	3	.000	48.282
11	6	6	.933	8	1.000	3.024	7	.000	151.165
12	7	2**	.000	8	.860	47.471	7	.098	51.805
13	6	6	.574	8	1.000	6.653	7	.000	206.707
14	2	2	.853	8	1.000	4.044	3	.000	33.666
15	2	2	.875	8	1.000	3.795	3	.000	37.724
16	2	2	.753	8	1.000	5.044	3	.000	43.902
17	8	8	.435	8	1.000	7.985	3	.000	113.642
18	8	8	.697	8	1.000	5.555	3	.000	81.424
19	8	8	.018	8	.997	18.406	2	.003	29.770
20	3	3	.189	8	1.000	11.239	2	.000	37.636
21	2	2	.887	8	1.000	3.653	3	.000	20.988
22	7	7	.912	8	1.000	3.330	3	.000	120.140
23	4	4	.631	8	1.000	6.147	1	.000	88.243
24	7	7	.548	8	1.000	6.892	3	.000	98.234
25	8	8	.002	8	1.000	23.776	3	.000	86.593
26	4	4	.503	8	1.000	7.315	8	.000	104.229
27	4	4	.266	8	1.000	9.981	3	.000	102.591
28	9	9	.621	8	1.000	6.234	3	.000	207.290
29	5	5	.251	8	1.000	10.206	3	.000	48.574
30	3	3	.611	8	1.000	6.324	2	.000	34.644
31	3	3	.576	8	1.000	6.643	2	.000	40.095
32	5	5	.246	8	1.000	10.284	3	.000	28.366
33	1	1	.098	8	1.000	13.418	3	.000	56.197
34	1	1	.618	8	1.000	6.262	3	.000	45.699
35	1	1	.319	8	1.000	9.285	2	.000	71.047
36	3	3	.454	8	1.000	7.788	2	.000	50.938
37	3	3	.191	8	1.000	11.198	2	.000	28.486
38	3	3	.516	8	1.000	7.194	2	.000	39.958
39	1	1	.577	8	1.000	6.627	2	.000	81.627
40	1	1	.574	8	1.000	6.657	2	.000	26.564
41	7	7	.865	8	1.000	3.915	2	.000	130.478
42	7	7	.562	8	1.000	6.769	2	.000	135.919
43	7	7	.902	8	1.000	3.459	2	.000	132.393
44	9	9	.713	8	1.000	5.409	3	.000	206.904
45	3	3	.331	8	.993	9.139	2	.007	18.968
46	3	3	.098	8	1.000	13.438	5	.000	32.129
47	7	7	.910	8	1.000	3.357	3	.000	108.681
48	5	5	.966	8	1.000	2.394	3	.000	71.126
49	7	7	.998	8	1.000	.999	3	.000	107.233
50	4	4	.003	8	1.000	23.244	5	.000	58.946

51	9	9	.000	8	1.000	31.529	3	.000	70.977
52	2	2	.670	8	.995	5.796	3	.005	16.450
53	3	3	1.000	8	1.000	.395	2	.000	23.636
54	3	3	.719	8	.998	5.355	2	.002	17.950
55	4	4	.638	8	1.000	6.086	8	.000	57.654
56	1	1	.967	8	1.000	2.381	2	.000	44.386
57	5	5	.442	8	1.000	7.908	3	.000	73.985
58	8	8	1.000	8	1.000	.617	3	.000	75.298
59	5	5	.514	8	1.000	7.215	3	.000	85.963
60	3	3	.960	8	1.000	2.538	2	.000	22.527
61	3	3	.854	8	1.000	4.039	2	.000	24.515
62	1	1	.784	8	1.000	4.747	2	.000	62.486
63	4	4	.465	8	1.000	7.685	2	.000	74.092
64	1	1	.574	8	1.000	6.659	2	.000	57.164
65	3	3	.973	8	1.000	2.226	2	.000	21.101
66	4	1**	.014	8	.668	19.230	4	.326	20.662
67	4	4	.547	8	1.000	6.907	1	.000	56.062
68	3	3	.675	8	1.000	5.752	2	.000	29.354
69	2	2	.508	8	.769	7.270	3	.231	9.671
70	3	3	.333	8	.935	9.116	2	.065	14.434
71	2	2	.030	8	1.000	16.966	3	.000	45.311
72	1	1	.029	8	1.000	17.114	5	.000	62.145
73	1	1	.868	8	1.000	3.873	2	.000	45.795
74	2	2	.572	8	.918	6.674	3	.082	11.494
75	6	6	.902	8	1.000	3.466	7	.000	202.989
76	2	2	.945	8	1.000	2.823	3	.000	26.134
77	1	1	.411	8	1.000	8.234	2	.000	64.576
78	2	2	.318	8	1.000	9.290	3	.000	24.521
79	3	3	.987	8	.999	1.776	2	.001	15.393
80	6	6	.006	8	1.000	21.540	7	.000	187.923
81	1	1	.994	8	1.000	1.395	3	.000	42.538
82	2	2	.398	8	1.000	8.376	3	.000	30.554
83	3	3	.876	8	1.000	3.784	2	.000	35.623
84	7	7	.647	8	1.000	6.001	3	.000	139.264
85	5	5	.989	8	1.000	1.708	3	.000	36.822
86	2	2	.483	8	1.000	7.506	3	.000	36.732
87	9	9	.293	8	1.000	9.623	2	.000	207.363
88	5	5	.996	8	1.000	1.258	3	.000	49.737
89	2	2	.675	8	1.000	5.754	3	.000	36.925
90	3	3	.996	8	1.000	1.238	2	.000	22.245
91	3	3	.987	8	1.000	1.770	2	.000	22.448
92	5	5	.913	8	1.000	3.317	3	.000	41.876
93	3	3	.386	8	.958	8.500	2	.042	14.766
94	6	6	.972	8	1.000	2.270	7	.000	210.613
95	3	3	.983	8	1.000	1.948	2	.000	29.159
96	3	3	.336	8	1.000	9.079	2	.000	29.387
97	5	5	.688	8	1.000	5.635	3	.000	50.577
98	1	1	.785	8	1.000	4.737	2	.000	66.692
99	2	2	.333	8	.997	9.113	3	.002	21.402
100	2	2	.463	8	.985	7.702	3	.015	16.095

101	5	5	.346	8	1.000	8.952	2	.000	41.638
102	5	5	.517	8	1.000	7.187	2	.000	39.906
103	5	5	.580	8	1.000	6.605	2	.000	69.169
104	1	1	.874	8	1.000	3.806	2	.000	67.646
105	1	1	.962	8	1.000	2.492	3	.000	38.530
106	5	5	.994	8	1.000	1.418	3	.000	38.590
107	3	3	.568	8	1.000	6.712	2	.000	24.020
108	3	3	.958	8	1.000	2.568	2	.000	25.778
109	4	4	.061	8	1.000	14.913	3	.000	37.547
110	2	2	.856	8	.991	4.017	3	.009	13.412
111	7	7	.468	8	1.000	7.650	2	.000	79.576
112	7	7	.000	8	1.000	49.635	1	.000	81.080
113	7	7	.011	8	1.000	19.936	6	.000	120.459
114	7	7	.000	8	1.000	38.755	6	.000	79.959
115	2	2	.165	8	1.000	11.709	3	.000	30.458
116	2	2	.916	8	1.000	3.268	3	.000	23.002
117	2	2	.751	8	.999	5.062	3	.001	18.459
118	1	1	.554	8	1.000	6.844	3	.000	42.790
119	1	1	.697	8	1.000	5.552	3	.000	71.128
120	3	3	.729	8	1.000	5.265	2	.000	44.185
121	8	8	.090	8	1.000	13.684	4	.000	31.040
122	5	5	.341	8	1.000	9.013	3	.000	80.349
123	6	6	.710	8	1.000	5.434	7	.000	173.519
124	3	3	1.000	8	1.000	.613	2	.000	17.889
125	6	6	.178	8	1.000	11.436	7	.000	272.067
126	9	9	.634	8	1.000	6.120	3	.000	215.493
127	3	3	.945	8	1.000	2.824	2	.000	19.194
128	3	3	.993	8	1.000	1.496	2	.000	21.194
129	1	1	.895	8	1.000	3.548	3	.000	63.716
130	6	6	.045	8	1.000	15.790	7	.000	242.552
131	8	8	.239	8	1.000	10.378	2	.000	118.844
132	4	4	.282	8	1.000	9.767	1	.000	48.084
133	4	4	.752	8	1.000	5.052	2	.000	57.486
134	8	8	.077	8	1.000	14.181	3	.000	39.286
135	8	8	.239	8	1.000	10.378	2	.000	118.844
136	3	3	.611	8	.944	6.322	2	.056	11.976
137	8	3**	.002	8	.695	24.323	8	.213	26.688
138	8	8	.012	8	.694	19.546	3	.168	22.382
139	2	2	.888	8	1.000	3.638	3	.000	37.898
140	7	7	.054	8	1.000	15.257	3	.000	79.301
141	1	1	1.000	8	1.000	.551	2	.000	49.050
142	5	5	.981	8	1.000	1.994	3	.000	41.233
143	3	3	.893	8	1.000	3.575	2	.000	32.660
144	1	1	.191	8	1.000	11.195	2	.000	95.064
145	7	7	.960	8	1.000	2.542	3	.000	119.994
146	7	7	.991	8	1.000	1.594	3	.000	114.463
147	7	7	.999	8	1.000	.737	3	.000	95.996
148	7	7	.433	8	1.000	8.006	2	.000	94.124
149	7	7	.999	8	1.000	.737	3	.000	95.996
150	9	9	.846	8	1.000	4.123	3	.000	197.781

151	7	7	.999	8	1.000	.737	3	.000	95.996
152	7	7	.987	8	1.000	1.765	3	.000	117.490
153	7	7	.834	8	1.000	4.253	3	.000	132.148
154	3	3	.954	8	1.000	2.660	2	.000	27.458
155	3	3	.672	8	.999	5.781	2	.001	18.831
156	9	9	.698	8	1.000	5.549	2	.000	186.779
157	1	1	.033	8	.999	16.766	5	.000	32.241
158	3	3	.678	8	1.000	5.728	2	.000	26.195
159	6	6	.961	8	1.000	2.506	7	.000	207.007
160	8	8	.003	8	1.000	23.425	5	.000	88.980
161	3	3	.670	8	1.000	5.793	2	.000	28.016
162	4	4	.995	8	1.000	1.370	3	.000	71.472
163	8	8	1.000	8	1.000	.617	3	.000	75.298
164	8	8	1.000	8	1.000	.617	3	.000	75.298
165	3	3	.924	8	1.000	3.160	2	.000	35.176
166	3	3	.675	8	.965	5.753	2	.035	12.401
167	1	1	.991	8	1.000	1.582	2	.000	35.122
168	3	3	.931	8	1.000	3.046	2	.000	31.656
169	1	1	.972	8	1.000	2.264	3	.000	45.857
170	1	1	.034	8	.758	16.670	2	.242	18.955
171	1	1	.952	8	1.000	2.704	2	.000	35.323
172	3	3	.006	8	1.000	21.462	2	.000	38.215
173	3	3	.930	8	1.000	3.068	2	.000	32.948
174	3	3	.694	8	1.000	5.578	2	.000	24.454
175	9	9	.833	8	1.000	4.263	3	.000	208.024
176	4	4	.296	8	1.000	9.573	8	.000	90.477
177	4	4	.077	8	1.000	14.187	2	.000	123.423
178	7	7	.026	8	1.000	17.421	1	.000	113.425
179	8	3**	.017	8	.866	18.608	4	.131	22.380
180	8	8	.376	8	1.000	8.614	3	.000	119.263
181	7	7	.969	8	1.000	2.336	2	.000	115.769
182	7	7	.120	8	1.000	12.781	2	.000	63.535
183	2	2	.967	8	1.000	2.383	3	.000	23.916
184	2	2	.967	8	1.000	2.383	3	.000	23.916
185	2	2	.919	8	.993	3.231	3	.007	13.134
186	6	6	.219	8	1.000	10.699	7	.000	148.553
187	6	6	.071	8	1.000	14.432	7	.000	144.422
188	6	6	.007	8	1.000	20.962	7	.000	121.832
189	1	1	.816	8	1.000	4.433	3	.000	55.908
190	3	3	.775	8	.988	4.837	2	.012	13.703
191	2	2	.895	8	1.000	3.550	3	.000	21.447
192	1	1	.044	8	.959	15.901	2	.029	22.899
193	5	5	.509	8	1.000	7.259	3	.000	74.000
194	3	3	.326	8	.985	9.202	5	.015	17.525
195	3	3	.015	8	.999	18.976	2	.001	33.149
196	2	2	.672	8	1.000	5.782	3	.000	27.171
197	4	4	.064	8	.994	14.769	2	.006	25.134
198	2	2	.184	8	1.000	11.324	3	.000	32.559
199	2	2	.826	8	1.000	4.329	3	.000	34.093
200	3	3	.222	8	.986	10.653	2	.013	19.239

201	5	5	.784	8	1.000	4.750	3	.000	59.571
202	1	1	.453	8	1.000	7.806	2	.000	25.306
203	5	5	.970	8	1.000	2.306	3	.000	58.109
204	5	5	.718	8	1.000	5.365	3	.000	64.571
205	3	3	.510	8	.953	7.250	2	.047	13.266
206	3	3	.593	8	1.000	6.483	2	.000	26.611
207	3	3	.868	8	1.000	3.875	2	.000	25.251
208	2	2	.682	8	1.000	5.693	3	.000	29.148
209	2	2	.700	8	1.000	5.526	3	.000	31.475
210	2	2	.769	8	1.000	4.891	3	.000	36.766
211	2	2	.910	8	1.000	3.359	3	.000	27.940
212	2	2	.355	8	.649	8.850	3	.351	10.083
213	2	2	.995	8	1.000	1.307	3	.000	18.965
214	2	2	.813	8	1.000	4.461	3	.000	24.332
215	2	2	.988	8	1.000	1.725	3	.000	19.570
216	3	3	.407	8	1.000	8.274	2	.000	26.090
217	9	3**	.000	8	.919	39.981	2	.061	45.409
218	2	2	.258	8	.994	10.098	1	.005	20.585
219	2	2	.603	8	1.000	6.394	3	.000	40.886
220	2	2	.665	8	1.000	5.839	3	.000	23.842
221	3	3	.782	8	1.000	4.772	2	.000	38.251
222	3	3	.999	8	1.000	.764	2	.000	26.884
223	3	3	.994	8	1.000	1.442	2	.000	17.907
224	2	2	.993	8	1.000	1.490	3	.000	33.159
225	2	2	.872	8	.966	3.826	3	.034	10.495
226	3	3	.904	8	.998	3.440	2	.002	16.378
227	2	2	.897	8	.999	3.529	3	.001	16.989
228	3	3	.497	8	1.000	7.372	2	.000	24.667
229	2	2	.828	8	.998	4.313	3	.002	16.456
230	9	9	.000	8	1.000	36.166	2	.000	60.451
231	5	5	.235	8	1.000	10.440	3	.000	49.383
232	6	6	.553	8	1.000	6.851	7	.000	197.704
233	3	3	.526	8	1.000	7.099	2	.000	46.296
234	3	3	.952	8	1.000	2.693	2	.000	23.217
235	2	2	.618	8	1.000	6.264	3	.000	22.975
236	1	1	.827	8	1.000	4.319	2	.000	69.264
237	3	3	.731	8	1.000	5.242	2	.000	27.935
238	3	3	.995	8	1.000	1.350	2	.000	23.870
239	2	2	.868	8	1.000	3.875	3	.000	33.956
240	2	2	.820	8	1.000	4.392	3	.000	31.464
241	1	1	.988	8	1.000	1.736	2	.000	55.212
242	3	3	.989	8	1.000	1.706	2	.000	17.451
243	9	9	.630	8	1.000	6.154	2	.000	195.002
244	5	5	.912	8	1.000	3.324	3	.000	48.467
245	2	2	.616	8	1.000	6.276	1	.000	45.327
246	5	5	.878	8	1.000	3.763	3	.000	55.614
247	4	4	.549	8	1.000	6.882	3	.000	55.166
248	1	1	.622	8	1.000	6.228	3	.000	29.459
249	1	1	.788	8	1.000	4.712	2	.000	77.141
250	1	1	.451	8	1.000	7.820	2	.000	26.806

251	2	2	.047	8	1.000	15.702	3	.000	50.450
252	5	5	.336	8	1.000	9.077	3	.000	26.498
253	7	7	.920	8	1.000	3.221	3	.000	106.025
254	4	4	.197	8	1.000	11.085	2	.000	115.159
255	8	8	.832	8	1.000	4.273	4	.000	99.982
256	8	8	.832	8	1.000	4.273	4	.000	99.982
257	8	8	.832	8	1.000	4.273	4	.000	99.982
258	8	8	.857	8	1.000	4.003	4	.000	98.184
259	8	8	.801	8	1.000	4.580	4	.000	94.136
260	8	8	.832	8	1.000	4.273	4	.000	99.982
261	1	1	.699	8	1.000	5.537	2	.000	28.790
262	4	4	.894	8	1.000	3.564	2	.000	54.739
263	4	4	.741	8	1.000	5.151	1	.000	56.770
264	5	5	.649	8	1.000	5.981	3	.000	64.931
265	3	3	.978	8	1.000	2.094	2	.000	25.645
266	2	2	.938	8	.998	2.940	3	.002	15.892
267	1	1	.346	8	1.000	8.961	3	.000	56.880
268	1	1	.548	8	1.000	6.899	2	.000	56.980
269	1	1	.948	8	1.000	2.765	2	.000	47.296
270	2	2	.918	8	1.000	3.241	3	.000	19.113
271	1	1	.897	8	1.000	3.528	2	.000	50.064
272	4	4	.953	8	1.000	2.672	1	.000	59.456
273	3	3	.818	8	1.000	4.414	2	.000	42.714
274	3	3	.667	8	1.000	5.820	2	.000	24.125
275	3	3	.583	8	1.000	6.576	2	.000	29.356
276	9	9	.008	8	1.000	20.778	3	.000	214.263
277	2	2	.039	8	1.000	16.274	3	.000	34.175

Table 66: DFA 8 Cluster Results Tests of Equality of Group Means for the HCA Within Groups Linkage Jaccard Coefficient Model

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.258	110.291	7	269	.000
Customer	.322	80.898	7	269	.000
TargetSpecific	.911	3.746	7	269	.001
Unassociated	.236	124.202	7	269	.000
Received	.518	35.794	7	269	.000
Introduced	.690	16.632	7	269	.000
Sought	.547	31.771	7	269	.000
WebsiteorOnlineAuction	.725	14.566	7	269	.000
Face2Face	.799	9.648	7	269	.000
Text	.952	1.946	7	269	.063
Phone	.734	13.915	7	269	.000
Seminar	.903	4.137	7	269	.000
InternetForum	.729	14.205	7	269	.000
InternetPopUp	.546	31.968	7	269	.000
Email	.727	14.408	7	269	.000
Post	.809	9.093	7	269	.000
Advertisement	.623	23.259	7	269	.000
Fax	.975	.969	7	269	.455
PrizeorMoney	.800	9.608	7	269	.000
HumanInteraction	.953	1.882	7	269	.073
FinancialReturn	.642	21.389	7	269	.000
Membership	.922	3.253	7	269	.002
AdviceorAssistance	.920	3.360	7	269	.002
Overpayment	.258	110.291	7	269	.000
Treatment	.895	4.515	7	269	.000
Employment	.552	31.182	7	269	.000
OpportunityForSelfOrOthers	.905	4.019	7	269	.000
Holiday	.964	1.435	7	269	.191
FinancialServices	.984	.625	7	269	.735
GoodLuck	.968	1.274	7	269	.263
Property	.980	.769	7	269	.614
Services	.958	1.704	7	269	.108
Merchandise	.709	15.773	7	269	.000
PartialPayment	.930	2.902	7	269	.006
Insight	.954	1.841	7	269	.080
Legal	.959	1.639	7	269	.125
FromFinancialInstitution	.717	15.191	7	269	.000
DetailUpdateorConfirmationRequired	.687	17.530	7	269	.000
GovernmentApproved	.953	1.897	7	269	.070
LoveAffectionConnection	.980	.791	7	269	.595
GovernmentAgency	.952	1.956	7	269	.061
LargeReturn	.511	36.834	7	269	.000
Effective	.925	3.120	7	269	.003
RefundAvailable	.988	.459	7	269	.864
FraudulentActivity	.684	17.750	7	269	.000
ShareTips	.824	8.220	7	269	.000
NoCreditCheckRequired	.966	1.342	7	269	.231
LittleorNoRisk	.823	8.291	7	269	.000
FromCorporateOrGovOfficial	.948	2.121	7	269	.042
QuickResponse	.947	2.151	7	269	.039
Confidentiality	.935	2.651	7	269	.011
PayupFrontCosts	.755	12.490	7	269	.000
ReceiveAndSendFunds	.841	7.287	7	269	.000
CallaPremiumNumber	.933	2.752	7	269	.009
TransferExcess	.333	76.909	7	269	.000
CompleteSaleoutsideofAuction	.930	2.902	7	269	.006
SendOntoOthers	.959	1.639	7	269	.124
RecruitOthers	.788	10.357	7	269	.000
SupplyPersonalInformation	.741	13.433	7	269	.000
SupplyBankAccDetails	.805	9.313	7	269	.000
Invest	.684	17.719	7	269	.000
MakeADonation	.918	3.411	7	269	.002
AlternativeShipment	.786	10.446	7	269	.000
Syntactic	.412	54.905	7	269	.000
Semantic	.335	76.299	7	269	.000
CompromisedWebsiteorFalseWebsite	.752	12.701	7	269	.000
DisguisedasInvoice	.949	2.063	7	269	.048
InferiorMerchandise	.943	2.321	7	269	.026
UseofFalsifiedForms	.620	23.568	7	269	.000
UseofParaphernalia	.862	6.177	7	269	.000
GoodsNeverSent	.821	8.373	7	269	.000
StoryBased	.875	5.471	7	269	.000
VerifiableStreetAddress	.985	.598	7	269	.758
LooksGenuine	.777	11.007	7	269	.000
ExploitLegitBusiness	.844	7.090	7	269	.000
Testimonials	.953	1.914	7	269	.068
RewardGreaterThanUpfrontCosts	.964	1.451	7	269	.185
FurtherContactbyEmailorPhone	.945	2.240	7	269	.031
PoliteBrokenEnglish	.966	1.330	7	269	.233
FinancialGain	.425	51.954	7	269	.000
Information	.532	33.803	7	269	.000
Participation	.783	10.645	7	269	.000

Table 67: DFA 8 Cluster Results Variable Failing Tolerance Testing for the HCA Within Groups Linkage Jaccard Coefficient Model

Variables Failing Tolerance Test[a]

	Within-Groups Variance	Tolerance	Minimum Tolerance
Overpayment	.008	.000	.000

All variables passing the tolerance criteria are entered simultaneously.

a. Minimum tolerance level is .001.

Table 68: DFA 8 Cluster Results Eigenvalues for the HCA Within Groups Linkage Jaccard Coefficient Model

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	7.465[a]	23.1	23.1	.939
2	5.991[a]	18.5	41.7	.926
3	5.358[a]	16.6	58.2	.918
4	4.355[a]	13.5	71.7	.902
5	4.047[a]	12.5	84.3	.895
6	2.819[a]	8.7	93.0	.859
7	2.268[a]	7.0	100.0	.833

a. First 7 canonical discriminant functions were used in the analysis.

Table 69: DFA 8 Cluster Results Function Significance Tests for the HCA Within Groups Linkage Jaccard Coefficient Model

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 7	.000	2720.380	567	.000
2 through 7	.000	2225.919	480	.000
3 through 7	.000	1775.748	395	.000
4 through 7	.003	1347.543	312	.000
5 through 7	.016	959.088	231	.000
6 through 7	.080	584.359	152	.000
7	.306	274.127	75	.000

Table 70: DFA 8 Cluster Results Predicted Groups Memberships for the HCA Within Groups Linkage Jaccard Coefficient Model

			Highest Group				Second Highest Group		
			P(D>d | G=g)						
Case Number	Actual Group	Predicted Group	p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid
1	1	1	.084	7	.975	12.550	2	.024	19.932
2	2	2	.999	7	1.000	.594	3	.000	18.741
3	3	3	.661	7	1.000	4.993	2	.000	42.695
4	1	1	.847	7	1.000	3.390	2	.000	53.217
5	2	2	.004	7	1.000	21.043	6	.000	64.796
6	4	4	.293	7	1.000	8.471	3	.000	36.406
7	2	2	.199	7	1.000	9.814	3	.000	53.434
8	5	5	.015	7	1.000	17.405	1	.000	38.735
9	3	3	.088	7	.925	12.394	5	.075	17.428
10	5	5	.901	7	1.000	2.819	3	.000	45.443
11	3	3	.998	7	1.000	.742	2	.000	25.054
12	6	2**	.272	7	.674	8.742	3	.326	10.190
13	3	3	.496	7	1.000	6.384	5	.000	36.011
14	2	2	.778	7	1.000	4.015	3	.000	34.393
15	2	2	.924	7	1.000	2.538	3	.000	36.269
16	2	2	.679	7	1.000	4.845	3	.000	44.594
17	7	7	.347	7	1.000	7.843	3	.000	113.671
18	7	7	.622	7	1.000	5.312	3	.000	82.383
19	7	7	.010	7	.997	18.437	2	.003	29.877
20	3	3	.474	7	1.000	6.581	2	.000	34.166
21	2	2	.821	7	1.000	3.635	3	.000	21.504
22	6	6	.907	7	1.000	2.755	3	.000	101.389
23	4	4	.544	7	1.000	5.960	1	.000	86.864
24	6	6	.787	7	1.000	3.941	3	.000	86.432
25	7	7	.002	7	1.000	23.110	3	.000	85.313
26	4	4	.404	7	1.000	7.247	7	.000	104.091
27	4	4	.200	7	1.000	9.803	3	.000	102.512
28	8	8	.720	7	1.000	4.502	3	.000	201.857
29	5	5	.191	7	1.000	9.966	3	.000	48.113
30	3	3	.563	7	1.000	5.800	2	.000	33.044
31	3	3	.783	7	1.000	3.970	2	.000	38.129
32	5	5	.208	7	1.000	9.666	3	.000	28.565
33	1	1	.234	7	1.000	9.268	3	.000	53.125
34	1	1	.864	7	1.000	3.221	3	.000	43.690
35	1	1	.234	7	1.000	9.272	2	.000	70.854
36	3	3	.464	7	1.000	6.668	2	.000	49.616
37	3	3	.136	7	1.000	11.060	2	.000	27.466
38	3	3	.728	7	1.000	4.443	2	.000	35.302
39	1	1	.467	7	1.000	6.638	2	.000	81.770
40	1	1	.604	7	1.000	5.458	2	.000	24.291
41	6	6	.801	7	1.000	3.811	2	.000	105.405
42	6	6	.488	7	1.000	6.453	2	.000	118.529
43	6	6	.840	7	1.000	3.451	3	.000	108.631
44	8	8	.611	7	1.000	5.404	3	.000	206.766
45	3	3	.226	7	.991	9.382	2	.009	18.889
46	3	3	.095	7	1.000	12.180	5	.000	31.368
47	6	6	.857	7	1.000	3.291	3	.000	84.900
48	5	5	.935	7	1.000	2.387	3	.000	71.656
49	6	6	.996	7	1.000	.913	3	.000	83.839
50	4	4	.002	7	1.000	23.075	5	.000	59.137

51	8	8	.000	7	1.000	31.575	3	.000	70.304
52	2	2	.797	7	.998	3.852	3	.002	16.174
53	3	3	1.000	7	1.000	.421	2	.000	23.202
54	3	3	.622	7	.997	5.313	2	.003	16.702
55	4	4	.534	7	1.000	6.054	7	.000	57.450
56	1	1	.969	7	1.000	1.831	2	.000	44.462
57	5	5	.349	7	1.000	7.812	3	.000	74.877
58	7	7	.999	7	1.000	.584	3	.000	75.606
59	5	5	.415	7	1.000	7.130	3	.000	86.883
60	3	3	.907	7	1.000	2.752	2	.000	22.012
61	3	3	.816	7	1.000	3.683	2	.000	24.028
62	1	1	.700	7	1.000	4.671	2	.000	62.201
63	4	4	.400	7	1.000	7.281	2	.000	72.654
64	1	1	.463	7	1.000	6.676	2	.000	57.218
65	3	3	.974	7	1.000	1.705	2	.000	20.641
66	4	1**	.009	7	.688	18.872	4	.307	20.488
67	4	4	.448	7	1.000	6.817	1	.000	55.315
68	3	3	.648	7	1.000	5.102	2	.000	27.605
69	2	2	.441	7	.844	6.884	3	.156	10.263
70	3	3	.369	7	.845	7.602	2	.155	10.994
71	2	2	.024	7	1.000	16.101	3	.000	44.285
72	1	1	.017	7	1.000	16.999	5	.000	61.254
73	1	1	.813	7	1.000	3.708	2	.000	45.600
74	2	2	.488	7	.941	6.456	3	.059	11.997
75	3	3	.762	7	1.000	4.153	2	.000	32.294
76	2	2	.903	7	1.000	2.797	3	.000	26.701
77	1	1	.376	7	1.000	7.524	2	.000	64.594
78	2	2	.447	7	1.000	6.828	3	.000	23.783
79	3	3	.955	7	.999	2.088	2	.001	15.136
80	3	3	.004	7	.998	20.836	2	.001	33.940
81	1	1	.989	7	1.000	1.263	3	.000	43.305
82	2	2	.324	7	1.000	8.100	3	.000	30.528
83	3	3	.918	7	1.000	2.616	2	.000	33.221
84	6	6	.569	7	1.000	5.752	3	.000	113.894
85	5	5	.976	7	1.000	1.664	3	.000	37.217
86	2	2	.378	7	1.000	7.513	3	.000	37.483
87	8	8	.222	7	1.000	9.445	2	.000	206.956
88	5	5	.991	7	1.000	1.187	3	.000	49.849
89	2	2	.573	7	1.000	5.720	3	.000	37.333
90	3	3	.991	7	1.000	1.217	2	.000	21.933
91	3	3	.982	7	1.000	1.522	2	.000	21.951
92	5	5	.863	7	1.000	3.225	3	.000	42.575
93	3	3	.308	7	.926	8.283	2	.074	13.342
94	3	3	.851	7	1.000	3.349	2	.000	35.000
95	3	3	.981	7	1.000	1.533	2	.000	27.744
96	3	3	.304	7	1.000	8.341	2	.000	28.852
97	5	5	.697	7	1.000	4.699	2	.000	51.195
98	1	1	.902	7	1.000	2.810	2	.000	63.496
99	2	2	.376	7	.999	7.527	3	.001	21.400
100	2	2	.369	7	.990	7.606	3	.010	16.816

101	5	5	.261	7	1.000	8.882	2	.000	41.717
102	5	5	.458	7	1.000	6.721	2	.000	39.213
103	5	5	.702	7	1.000	4.653	2	.000	66.872
104	1	1	.803	7	1.000	3.799	2	.000	67.731
105	1	1	.931	7	1.000	2.442	3	.000	39.346
106	5	5	.986	7	1.000	1.372	3	.000	39.228
107	3	3	.535	7	1.000	6.039	2	.000	21.921
108	3	3	.982	7	1.000	1.498	2	.000	23.359
109	4	4	.061	7	1.000	13.487	3	.000	33.112
110	2	2	.824	7	.995	3.602	3	.005	14.023
111	6	6	.960	7	1.000	2.003	3	.000	72.346
112	6	1**	.004	7	.782	20.724	3	.214	23.314
113	6	6	.977	7	1.000	1.645	3	.000	89.723
114	6	6	.982	7	1.000	1.496	3	.000	66.943
115	2	2	.285	7	1.000	8.569	1	.000	26.080
116	2	2	.943	7	1.000	2.276	3	.000	23.450
117	2	2	.780	7	.999	4.000	3	.001	18.836
118	1	1	.576	7	1.000	5.696	3	.000	42.388
119	1	1	.639	7	1.000	5.170	3	.000	71.763
120	3	3	.670	7	1.000	4.914	2	.000	43.594
121	7	7	.061	7	1.000	13.485	4	.000	31.114
122	5	5	.250	7	1.000	9.040	3	.000	80.856
123	3	3	.609	7	1.000	5.419	2	.000	44.986
124	3	3	.997	7	1.000	.800	2	.000	17.605
125	3	3	.266	7	1.000	8.815	2	.000	50.896
126	8	8	.530	7	1.000	6.082	3	.000	215.732
127	3	3	.921	7	1.000	2.585	2	.000	18.709
128	3	3	.983	7	1.000	1.470	2	.000	20.805
129	1	1	.829	7	1.000	3.554	3	.000	64.656
130	3	3	.048	7	1.000	14.200	2	.000	44.336
131	7	7	.187	7	1.000	10.025	2	.000	119.128
132	4	4	.218	7	1.000	9.511	1	.000	47.755
133	4	4	.669	7	1.000	4.927	2	.000	57.581
134	7	7	.086	7	1.000	12.463	3	.000	39.435
135	7	7	.187	7	1.000	10.025	2	.000	119.128
136	3	3	.507	7	.894	6.286	2	.106	10.555
137	7	3**	.001	7	.584	24.858	7	.267	26.427
138	7	7	.007	7	.680	19.530	3	.172	22.273
139	2	2	.831	7	1.000	3.544	3	.000	38.611
140	6	6	.243	7	1.000	9.142	3	.000	72.225
141	1	1	.999	7	1.000	.542	2	.000	48.891
142	5	5	.964	7	1.000	1.918	3	.000	41.918
143	3	3	.974	7	1.000	1.710	2	.000	31.294
144	1	1	.152	7	1.000	10.703	2	.000	93.958
145	6	6	.961	7	1.000	1.975	3	.000	100.750
146	6	6	.980	7	1.000	1.554	3	.000	87.130
147	6	6	.999	7	1.000	.662	3	.000	73.242
148	6	6	.354	7	1.000	7.757	2	.000	76.710
149	6	6	.999	7	1.000	.662	3	.000	73.242
150	8	8	.767	7	1.000	4.114	6	.000	195.470

151	6	6	.999	7	1.000	.662	3	.000	73.242
152	6	6	.974	7	1.000	1.707	3	.000	92.058
153	6	6	.788	7	1.000	3.930	3	.000	101.387
154	3	3	.907	7	1.000	2.755	2	.000	26.975
155	3	3	.560	7	.998	5.826	2	.002	18.641
156	8	8	.610	7	1.000	5.412	2	.000	185.909
157	1	1	.021	7	.999	16.508	5	.001	30.740
158	3	3	.512	7	1.000	6.242	2	.000	26.033
159	3	3	.903	7	1.000	2.794	2	.000	29.261
160	7	7	.002	7	1.000	22.889	5	.000	88.686
161	3	3	.856	7	1.000	3.297	2	.000	23.638
162	4	4	.995	7	1.000	.969	3	.000	72.182
163	7	7	.999	7	1.000	.584	3	.000	75.606
164	7	7	.999	7	1.000	.584	3	.000	75.606
165	3	3	.921	7	1.000	2.580	2	.000	33.754
166	3	3	.534	7	.958	6.052	2	.042	12.296
167	1	1	.989	7	1.000	1.290	2	.000	35.231
168	3	3	.936	7	1.000	2.373	2	.000	30.032
169	1	1	.956	7	1.000	2.073	3	.000	46.582
170	1	1	.022	7	.676	16.307	2	.324	17.777
171	1	1	.913	7	1.000	2.682	2	.000	35.275
172	3	3	.961	7	1.000	1.971	2	.000	21.692
173	3	3	.982	7	1.000	1.510	2	.000	29.948
174	3	3	.561	7	1.000	5.819	2	.000	24.292
175	8	8	.754	7	1.000	4.221	6	.000	207.330
176	4	4	.254	7	1.000	8.984	7	.000	90.644
177	4	4	.048	7	1.000	14.189	2	.000	123.156
178	6	6	.038	7	1.000	14.846	1	.000	100.356
179	7	3**	.014	7	.913	17.611	4	.085	22.371
180	7	7	.350	7	1.000	7.805	2	.000	120.154
181	6	6	.956	7	1.000	2.073	2	.000	98.516
182	6	6	.293	7	1.000	8.468	2	.000	56.691
183	2	2	.962	7	1.000	1.969	3	.000	23.586
184	2	2	.962	7	1.000	1.969	3	.000	23.586
185	2	2	.898	7	.996	2.859	3	.004	13.740
186	3	3	.052	7	1.000	13.954	2	.000	45.626
187	3	3	.017	7	.999	17.065	6	.001	30.282
188	3	3	.945	7	1.000	2.249	2	.000	18.547
189	1	1	.935	7	1.000	2.398	3	.000	54.636
190	3	3	.707	7	.987	4.610	2	.013	13.219
191	2	2	.865	7	1.000	3.213	3	.000	22.136
192	1	1	.034	7	.971	15.161	2	.021	22.784
193	5	5	.409	7	1.000	7.190	3	.000	74.064
194	3	3	.218	7	.980	9.518	5	.020	17.260
195	3	3	.054	7	1.000	13.843	2	.000	29.249
196	2	2	.578	7	1.000	5.674	3	.000	27.532
197	4	4	.046	7	.996	14.289	2	.004	25.220
198	2	2	.127	7	1.000	11.267	3	.000	33.373
199	2	2	.757	7	1.000	4.194	3	.000	34.928
200	3	3	.229	7	.989	9.350	2	.011	18.270

201	5	5	.696	7	1.000	4.704	3	.000	60.039
202	1	1	.395	7	1.000	7.336	2	.000	24.008
203	5	5	.980	7	1.000	1.574	3	.000	58.947
204	5	5	.680	7	1.000	4.839	3	.000	63.688
205	3	3	.382	7	.916	7.466	2	.084	12.235
206	3	3	.961	7	1.000	1.985	2	.000	19.857
207	3	3	.903	7	1.000	2.803	2	.000	23.008
208	2	2	.605	7	1.000	5.452	3	.000	29.933
209	2	2	.604	7	1.000	5.456	3	.000	32.031
210	2	2	.693	7	1.000	4.728	3	.000	37.109
211	2	2	.849	7	1.000	3.368	3	.000	28.585
212	2	2	.435	7	.795	6.941	3	.205	9.645
213	2	2	.990	7	1.000	1.232	3	.000	19.275
214	2	2	.729	7	1.000	4.430	3	.000	24.945
215	2	2	.981	7	1.000	1.532	3	.000	19.616
216	3	3	.329	7	1.000	8.037	2	.000	24.753
217	8	3**	.000	7	.952	37.054	2	.043	43.266
218	2	2	.185	7	.994	10.065	1	.005	20.500
219	2	2	.526	7	1.000	6.121	3	.000	40.860
220	2	2	.594	7	1.000	5.542	3	.000	24.620
221	3	3	.826	7	1.000	3.590	2	.000	35.913
222	3	3	.999	7	1.000	.673	2	.000	26.218
223	3	3	.985	7	1.000	1.412	2	.000	16.991
224	2	2	.983	7	1.000	1.472	3	.000	33.787
225	2	2	.801	7	.974	3.810	3	.026	11.020
226	3	3	.837	7	.998	3.484	2	.002	15.690
227	2	2	.918	7	.999	2.616	3	.001	17.408
228	3	3	.401	7	1.000	7.272	2	.000	23.426
229	2	2	.953	7	.997	2.119	3	.003	13.606
230	8	8	.000	7	1.000	35.973	2	.000	60.304
231	5	5	.175	7	1.000	10.239	3	.000	50.069
232	3	3	.555	7	1.000	5.872	2	.000	25.084
233	3	3	.642	7	1.000	5.149	2	.000	43.011
234	3	3	.943	7	1.000	2.274	2	.000	21.799
235	2	2	.509	7	1.000	6.262	3	.000	23.445
236	1	1	.791	7	1.000	3.898	2	.000	68.250
237	3	3	.657	7	1.000	5.023	2	.000	27.664
238	3	3	.985	7	1.000	1.409	2	.000	23.621
239	2	2	.838	7	1.000	3.476	3	.000	33.727
240	2	2	.806	7	1.000	3.770	3	.000	30.890
241	1	1	.974	7	1.000	1.715	2	.000	55.000
242	3	3	.980	7	1.000	1.572	2	.000	17.064
243	8	8	.578	7	1.000	5.674	2	.000	195.644
244	5	5	.866	7	1.000	3.198	3	.000	48.304
245	2	2	.533	7	1.000	6.054	1	.000	45.373
246	5	5	.808	7	1.000	3.751	3	.000	55.916
247	4	4	.511	7	1.000	6.249	3	.000	55.819
248	1	1	.522	7	1.000	6.155	3	.000	30.196
249	1	1	.775	7	1.000	4.044	2	.000	75.728
250	1	1	.366	7	1.000	7.629	2	.000	26.873

251	2	2	.083	7	1.000	12.595	3	.000	46.522
252	5	5	.326	7	1.000	8.079	3	.000	27.091
253	6	6	.931	7	1.000	2.449	3	.000	88.832
254	4	4	.136	7	1.000	11.056	2	.000	114.751
255	7	7	.763	7	1.000	4.143	3	.000	100.129
256	7	7	.763	7	1.000	4.143	3	.000	100.129
257	7	7	.763	7	1.000	4.143	3	.000	100.129
258	7	7	.778	7	1.000	4.012	4	.000	98.355
259	7	7	.712	7	1.000	4.572	4	.000	94.345
260	7	7	.763	7	1.000	4.143	3	.000	100.129
261	1	1	.644	7	1.000	5.134	2	.000	27.621
262	4	4	.831	7	1.000	3.539	2	.000	54.618
263	4	4	.644	7	1.000	5.134	1	.000	55.699
264	5	5	.541	7	1.000	5.985	3	.000	65.389
265	3	3	.999	7	1.000	.611	2	.000	22.556
266	2	2	.892	7	.999	2.929	3	.001	16.330
267	1	1	.264	7	1.000	8.846	3	.000	57.674
268	1	1	.454	7	1.000	6.764	2	.000	56.646
269	1	1	.929	7	1.000	2.476	2	.000	47.388
270	2	2	.885	7	1.000	3.005	3	.000	19.850
271	1	1	.849	7	1.000	3.373	2	.000	50.180
272	4	4	.915	7	1.000	2.659	1	.000	58.094
273	3	3	.806	7	1.000	3.772	2	.000	41.626
274	3	3	.536	7	1.000	6.031	2	.000	23.848
275	3	3	.467	7	1.000	6.641	2	.000	28.912
276	8	8	.005	7	1.000	20.512	3	.000	214.413
277	2	2	.041	7	1.000	14.644	3	.000	32.036

Table 71: DFA 7 Cluster Results Tests of Equality of Group Means for the HCA Furthest Neighbour Jaccard Coefficient Model with Insignificant Features Removed

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Seller	.904	4.758	6	270	.000
Customer	.268	123.176	6	270	.000
TargetSpecific	.946	2.552	6	270	.020
Unassociated	.411	64.366	6	270	.000
Received	.559	35.474	6	270	.000
Introduced	.786	12.242	6	270	.000
Sought	.667	22.466	6	270	.000
WebsiteorOnlineAuction	.824	9.589	6	270	.000
Face2Face	.853	7.736	6	270	.000
Text	.820	9.885	6	270	.000
Phone	.935	3.124	6	270	.006
Seminar	.910	4.465	6	270	.000
InternetForum	.837	8.753	6	270	.000
InternetPopUp	.791	11.908	6	270	.000
Email	.773	13.185	6	270	.000
Post	.807	10.787	6	270	.000
Advertisement	.753	14.752	6	270	.000
Fax	.955	2.142	6	270	.049
PrizeorMoney	.638	25.566	6	270	.000
FinancialReturn	.626	26.842	6	270	.000
Membership	.860	7.307	6	270	.000
AdviceorAssistance	.908	4.584	6	270	.000
Treatment	.895	5.300	6	270	.000
Employment	.156	243.588	6	270	.000
OpportunityForSelfOrOthers	.793	11.764	6	270	.000
Holiday	.932	3.259	6	270	.004
FinancialServices	.946	2.550	6	270	.020
Property	.942	2.747	6	270	.013
Services	.927	3.536	6	270	.002
Merchandise	.672	21.982	6	270	.000
PartialPayment	.923	3.736	6	270	.001
Legal	.932	3.301	6	270	.004
FromFinancialInstitution	.874	6.511	6	270	.000
DetailUpdateorConfirmationRequired	.709	18.475	6	270	.000
GovernmentApproved	.951	2.332	6	270	.033
LargeReturn	.617	27.938	6	270	.000
Effective	.900	5.007	6	270	.000
FraudulentActivity	.882	6.009	6	270	.000
ShareTips	.932	3.258	6	270	.004
LittleorNoRisk	.878	6.253	6	270	.000
QuickResponse	.949	2.426	6	270	.027
Confidentiality	.910	4.436	6	270	.000
PayupFrontCosts	.758	14.388	6	270	.000
ReceiveAndSendFunds	.778	12.822	6	270	.000
CallaPremiumNumber	.740	15.795	6	270	.000
TransferExcess	.916	4.125	6	270	.001
CompleteSaleoutsideofAuction	.923	3.736	6	270	.001
RecruitOthers	.735	16.187	6	270	.000
SupplyPersonalInformation	.764	13.921	6	270	.000
SupplyBankAccDetails	.797	11.468	6	270	.000
Invest	.688	20.437	6	270	.000
MakeADonation	.912	4.336	6	270	.000
AlternativeShipment	.767	13.700	6	270	.000
Syntactic	.659	23.246	6	270	.000
Semantic	.680	21.219	6	270	.000
CompromisedWebsiteorFalseWebsite	.791	11.905	6	270	.000
InferiorMerchandise	.919	3.975	6	270	.001
UseofFalsifiedForms	.853	7.777	6	270	.000
UseofParaphernalia	.854	7.708	6	270	.000
GoodsNeverSent	.804	10.967	6	270	.000
StoryBased	.809	10.598	6	270	.000
LooksGenuine	.880	6.125	6	270	.000
ExploitLegitBusiness	.924	3.727	6	270	.001
Testimonials	.921	3.870	6	270	.001
RewardGreaterThanUpfrontCosts	.945	2.631	6	270	.017
FinancialGain	.326	92.958	6	270	.000
Information	.498	45.412	6	270	.000
Participation	.653	23.893	6	270	.000

Table 72: DFA 7 Cluster Results Eigenvalues for the HCA Furthest Neighbour Jaccard Coefficient Model with Insignificant Features Removed

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	10.205[a]	32.5	32.5	.954
2	8.695[a]	27.7	60.3	.947
3	5.732[a]	18.3	78.6	.923
4	3.237[a]	10.3	88.9	.874
5	1.927[a]	6.1	95.0	.811
6	1.560[a]	5.0	100.0	.781

Table 73: DFA 7 Cluster Results Function Significance Tests for the HCA Furthest Neighbour Jaccard Coefficient Model with Insignificant Features Removed

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 6	.000	2397.577	408	.000
2 through 6	.000	1821.273	335	.000
3 through 6	.005	1279.497	264	.000
4 through 6	.031	824.709	195	.000
5 through 6	.133	480.349	128	.000
6	.391	224.213	63	.000

Table 74: DFA 7 Cluster Results Predicted Groups Memberships for the HCA Furthest Neighbour Jaccard Coefficient Model with Insignificant Features Removed

		Highest Group					Second Highest Group		
			P(D>d | G=g)						
Case Number	Actual Group	Predicted Group	p	df	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid	Group	P(G=g | D=d)	Squared Mahalanobis Distance to Centroid
1	1	1	.281	6	.834	7.453	7	.161	10.741
2	2	2	.887	6	1.000	2.329	1	.000	31.508
3	2	2	.464	6	1.000	5.647	1	.000	36.775
4	1	1	.503	6	1.000	5.322	2	.000	25.420
5	1	1	.004	6	1.000	19.193	2	.000	64.326
6	2	2	.134	6	1.000	9.793	4	.000	35.347
7	1	1	.289	6	1.000	7.358	7	.000	29.285
8	3	3	.035	6	1.000	13.580	7	.000	97.453
9	3	3	.810	6	1.000	2.993	1	.000	117.574
10	3	3	.955	6	1.000	1.568	1	.000	146.706
11	1	1	.210	6	1.000	8.407	2	.000	33.762
12	1	1	.678	6	1.000	3.989	7	.000	23.597
13	1	1	.915	6	1.000	2.052	2	.000	27.792
14	4	4	.026	6	1.000	14.392	2	.000	72.555
15	4	4	.075	6	1.000	11.478	1	.000	79.822
16	4	4	.695	6	1.000	3.866	1	.000	44.734
17	5	5	.917	6	1.000	2.034	2	.000	58.596
18	4	4	.245	6	1.000	7.907	5	.000	45.043
19	5	1**	.047	6	.560	12.741	5	.434	13.251
20	2	2	.493	6	1.000	5.402	1	.000	28.308
21	1	1	.812	6	1.000	2.978	2	.000	18.915
22	6	6	.805	6	1.000	3.033	1	.000	141.371
23	7	7	.024	6	1.000	14.549	1	.000	69.850
24	6	6	.078	6	1.000	11.338	5	.000	113.114
25	5	5	.027	6	1.000	14.237	1	.000	30.935
26	5	5	.902	6	1.000	2.189	1	.000	68.011
27	5	5	.608	6	1.000	4.511	2	.000	77.029
28	1	1	.695	6	1.000	3.862	2	.000	35.179
29	3	3	.618	6	1.000	4.433	5	.000	136.545
30	2	2	.448	6	1.000	5.783	1	.000	33.521
31	2	2	.579	6	1.000	4.729	1	.000	42.969
32	4	4	.106	6	.814	10.484	2	.186	13.442
33	7	7	.433	6	.991	5.909	1	.009	15.293
34	7	7	.892	6	1.000	2.284	1	.000	40.336
35	7	7	.971	6	1.000	1.310	1	.000	21.373
36	2	2	.810	6	1.000	2.988	1	.000	28.219
37	2	2	.895	6	1.000	2.257	1	.000	24.280
38	2	2	.456	6	1.000	5.712	1	.000	31.739
39	7	7	.440	6	1.000	5.852	1	.000	56.235
40	1	1	.941	6	1.000	1.758	7	.000	24.704
41	6	6	.972	6	1.000	1.297	1	.000	128.245
42	6	6	.261	6	1.000	7.698	2	.000	185.468
43	6	6	.886	6	1.000	2.337	1	.000	156.228
44	1	1	.857	6	1.000	2.605	2	.000	31.692
45	5	1**	.025	6	.636	14.432	5	.322	15.790
46	5	2**	.029	6	.976	14.081	5	.017	22.207
47	6	6	.697	6	1.000	3.852	2	.000	132.041
48	3	3	.976	6	1.000	1.218	1	.000	158.095
49	6	6	.997	6	1.000	.555	1	.000	134.064
50	5	3**	.000	6	.922	49.468	5	.078	54.409

51	1	1	.066	6	1.000	11.828	5	.000	33.478
52	7	7	.117	6	1.000	10.189	4	.000	31.664
53	2	2	.996	6	1.000	.617	1	.000	24.442
54	2	2	.771	6	.989	3.296	1	.011	12.370
55	5	5	.994	6	1.000	.724	1	.000	55.671
56	7	7	.470	6	.874	5.599	1	.126	9.465
57	3	3	.219	6	1.000	8.276	5	.000	146.028
58	5	5	.977	6	1.000	1.193	1	.000	59.490
59	3	3	.707	6	1.000	3.775	4	.000	142.772
60	2	2	.535	6	.854	5.072	1	.146	8.606
61	2	2	.966	6	1.000	1.405	1	.000	18.987
62	1	1	.370	6	.874	6.495	7	.126	10.369
63	5	5	.611	6	1.000	4.489	1	.000	52.496
64	1	7**	.090	6	.610	10.941	1	.390	11.838
65	2	2	.999	6	1.000	.339	1	.000	18.562
66	1	1	.868	6	1.000	2.504	2	.000	17.965
67	5	5	.084	6	1.000	11.142	1	.000	45.154
68	2	2	.207	6	1.000	8.456	1	.000	47.633
69	2	2	.632	6	.999	4.330	1	.001	17.567
70	1	1	.895	6	.999	2.254	2	.001	15.921
71	2	2	.003	6	1.000	19.781	1	.000	43.632
72	7	3**	.000	6	.950	60.164	7	.050	66.036
73	1	1	.496	6	.995	5.379	7	.005	15.818
74	2	2	.738	6	1.000	3.547	1	.000	21.709
75	1	1	.731	6	1.000	3.600	7	.000	30.770
76	4	4	.343	6	1.000	6.768	2	.000	53.871
77	7	7	.053	6	1.000	12.446	1	.000	29.269
78	2	2	.049	6	1.000	12.667	1	.000	47.399
79	2	2	.767	6	.998	3.328	1	.002	16.328
80	1	1	.068	6	1.000	11.755	7	.000	32.412
81	1	1	.620	6	.985	4.424	7	.015	12.767
82	1	2**	.008	6	.614	17.342	1	.386	18.274
83	2	2	.597	6	1.000	4.591	1	.000	22.362
84	6	6	.724	6	1.000	3.652	2	.000	158.560
85	3	3	.944	6	1.000	1.708	2	.000	117.337
86	2	2	.090	6	1.000	10.941	1	.000	32.172
87	1	1	.499	6	1.000	5.355	2	.000	33.303
88	3	3	.912	6	1.000	2.081	1	.000	146.954
89	2	2	.518	6	.998	5.200	1	.002	17.828
90	2	2	.992	6	1.000	.815	1	.000	20.947
91	2	2	.767	6	.998	3.328	1	.002	16.328
92	3	3	.993	6	1.000	.749	1	.000	123.777
93	2	2	.958	6	1.000	1.527	1	.000	22.036
94	1	1	.798	6	1.000	3.086	2	.000	31.263
95	2	2	.862	6	1.000	2.553	1	.000	18.312
96	1	1	.387	6	.925	6.329	2	.075	11.360
97	4	4	.652	6	1.000	4.184	1	.000	32.413
98	7	7	.499	6	1.000	5.356	1	.000	45.949
99	2	2	.372	6	.854	6.473	1	.146	10.001
100	4	4	.736	6	.999	3.558	2	.001	16.633

101	4	4	.855	6	1.000	2.619	1	.000	28.248
102	4	4	.926	6	1.000	1.934	1	.000	24.554
103	4	4	.562	6	1.000	4.859	1	.000	40.595
104	1	1	.684	6	1.000	3.945	7	.000	19.582
105	7	7	.323	6	1.000	6.979	2	.000	40.269
106	4	4	.537	6	.993	5.054	2	.007	15.050
107	2	2	.395	6	.999	6.259	1	.001	19.313
108	2	2	.560	6	1.000	4.873	1	.000	37.171
109	5	5	.540	6	1.000	5.028	4	.000	39.656
110	2	2	.837	6	1.000	2.770	1	.000	19.684
111	6	6	.693	6	1.000	3.879	1	.000	93.985
112	1	1	.921	6	1.000	1.988	2	.000	25.847
113	6	6	.828	6	1.000	2.845	1	.000	110.226
114	6	6	.068	6	1.000	11.729	1	.000	82.863
115	7	7	.965	6	1.000	1.416	1	.000	24.757
116	1	1	.736	6	.999	3.562	2	.001	16.710
117	1	1	.408	6	.998	6.141	2	.001	19.457
118	7	7	.693	6	.997	3.880	1	.003	15.556
119	7	7	.820	6	1.000	2.914	1	.000	38.454
120	2	2	.337	6	1.000	6.833	7	.000	24.541
121	5	5	.997	6	1.000	.533	1	.000	55.759
122	3	3	.449	6	1.000	5.772	4	.000	145.957
123	1	1	.754	6	1.000	3.421	2	.000	26.365
124	2	2	.896	6	1.000	2.245	4	.000	23.156
125	1	1	.955	6	1.000	1.562	2	.000	27.512
126	1	1	.942	6	1.000	1.735	2	.000	24.449
127	2	2	.674	6	.981	4.023	1	.019	11.915
128	2	2	.483	6	.970	5.488	4	.016	13.702
129	7	7	.955	6	1.000	1.567	1	.000	30.305
130	1	1	.428	6	1.000	5.961	2	.000	24.766
131	5	5	.977	6	1.000	1.188	1	.000	58.284
132	5	5	.740	6	1.000	3.530	1	.000	38.803
133	5	5	.937	6	1.000	1.805	1	.000	49.946
134	5	5	.608	6	1.000	4.510	2	.000	31.722
135	5	5	.977	6	1.000	1.188	1	.000	58.284
136	2	2	.091	6	1.000	10.923	1	.000	55.636
137	2	2	.146	6	.996	9.522	1	.002	21.845
138	2	2	.487	6	.914	5.458	1	.086	10.188
139	2	2	.536	6	1.000	5.058	1	.000	23.703
140	6	6	.162	6	1.000	9.204	2	.000	79.153
141	1	1	.920	6	.999	1.993	7	.001	15.891
142	3	3	.200	6	1.000	8.551	1	.000	131.279
143	2	2	.654	6	1.000	4.171	1	.000	29.965
144	7	7	.080	6	1.000	11.284	1	.000	68.350
145	6	6	.666	6	1.000	4.077	2	.000	164.528
146	6	6	.926	6	1.000	1.926	1	.000	111.297
147	6	6	.976	6	1.000	1.225	2	.000	112.510
148	6	6	.976	6	1.000	1.225	2	.000	112.510
149	6	6	.976	6	1.000	1.225	2	.000	112.510
150	1	1	.986	6	1.000	.992	2	.000	25.876

151	6	6	.976	6	1.000	1.225	2	.000	112.510
152	6	6	.987	6	1.000	.960	2	.000	139.947
153	6	6	.858	6	1.000	2.591	2	.000	159.601
154	2	2	.820	6	1.000	2.907	1	.000	24.394
155	2	2	.005	6	.601	18.363	5	.398	19.187
156	1	1	.748	6	1.000	3.466	7	.000	31.887
157	3	3	.254	6	1.000	7.792	1	.000	138.431
158	3	3	.371	6	1.000	6.481	7	.000	112.385
159	1	1	.881	6	1.000	2.381	2	.000	27.386
160	5	5	.204	6	1.000	8.495	4	.000	51.385
161	7	7	.180	6	.973	8.880	2	.027	16.078
162	5	5	.876	6	1.000	2.433	1	.000	33.516
163	5	5	.977	6	1.000	1.193	1	.000	59.490
164	5	5	.977	6	1.000	1.193	1	.000	59.490
165	2	2	.681	6	1.000	3.965	1	.000	32.756
166	2	2	.063	6	.738	11.974	4	.262	14.047
167	7	7	.850	6	1.000	2.659	1	.000	21.392
168	2	2	.904	6	1.000	2.162	1	.000	28.641
169	7	7	.287	6	.830	7.387	1	.168	10.585
170	1	1	.892	6	1.000	2.285	2	.000	28.613
171	1	1	.938	6	1.000	1.790	2	.000	24.316
172	1	1	.803	6	1.000	3.046	2	.000	18.403
173	2	2	.305	6	1.000	7.174	1	.000	50.123
174	2	2	.173	6	1.000	9.008	1	.000	46.192
175	1	1	.952	6	1.000	1.602	2	.000	32.952
176	5	5	.698	6	1.000	3.844	1	.000	64.138
177	5	5	.533	6	1.000	5.086	1	.000	75.934
178	6	6	.034	6	1.000	13.620	5	.000	115.114
179	5	5	.846	6	1.000	2.692	1	.000	32.106
180	5	5	.377	6	1.000	6.428	1	.000	79.535
181	6	6	.998	6	1.000	.511	2	.000	137.683
182	6	6	.005	6	1.000	18.443	4	.000	76.791
183	4	4	.866	6	1.000	2.525	1	.000	22.400
184	4	4	.866	6	1.000	2.525	1	.000	22.400
185	4	4	.464	6	.994	5.647	2	.006	15.874
186	1	1	.406	6	1.000	6.156	7	.000	30.296
187	1	1	.001	6	1.000	22.434	7	.000	50.204
188	1	1	.784	6	.999	3.197	2	.001	16.290
189	1	1	.981	6	1.000	1.105	2	.000	17.965
190	2	2	.168	6	.996	9.096	4	.004	19.919
191	7	7	.082	6	.559	11.220	2	.383	11.975
192	1	1	.178	6	.993	8.919	2	.006	19.075
193	3	3	.133	6	1.000	9.799	5	.000	151.294
194	3	3	.715	6	1.000	3.715	2	.000	109.349
195	1	1	.114	6	.997	10.270	7	.003	22.111
196	1	1	.839	6	1.000	2.756	2	.000	20.027
197	1	1	.193	6	.997	8.668	4	.002	21.121
198	4	4	.617	6	1.000	4.440	2	.000	41.741
199	4	4	.415	6	.999	6.077	1	.001	19.343
200	2	2	.007	6	1.000	17.700	1	.000	34.727

201	3	3	.735	6	1.000	3.566	2	.000	126.622
202	7	3**	.000	6	.996	35.793	7	.004	46.996
203	3	3	.734	6	1.000	3.572	1	.000	123.208
204	3	3	.641	6	1.000	4.263	4	.000	120.899
205	1	1	.851	6	1.000	2.650	7	.000	20.215
206	2	2	.033	6	.997	13.742	4	.003	25.577
207	2	1**	.470	6	.524	5.594	2	.476	5.788
208	2	2	.694	6	1.000	3.875	1	.000	19.432
209	1	1	.762	6	1.000	3.365	2	.000	18.653
210	1	1	.778	6	1.000	3.239	2	.000	22.273
211	1	1	.706	6	.998	3.780	2	.001	17.470
212	1	2**	.707	6	.924	3.775	1	.076	8.785
213	1	1	.924	6	1.000	1.954	2	.000	21.127
214	1	1	.881	6	1.000	2.386	2	.000	22.394
215	1	1	.433	6	1.000	5.914	7	.000	21.764
216	1	1	.873	6	.998	2.456	2	.002	14.675
217	1	1	.123	6	1.000	10.028	2	.000	30.731
218	1	1	.895	6	1.000	2.252	2	.000	26.143
219	1	1	.931	6	1.000	1.873	2	.000	29.236
220	1	2**	.182	6	.769	8.856	1	.230	11.267
221	2	2	.947	6	1.000	1.673	4	.000	28.600
222	2	2	.997	6	1.000	.582	1	.000	20.394
223	2	2	.987	6	1.000	.962	1	.000	17.543
224	1	1	.209	6	.691	8.423	2	.309	10.029
225	1	1	.549	6	.996	4.959	2	.004	16.220
226	2	2	.849	6	1.000	2.666	4	.000	25.953
227	2	1**	.845	6	.992	2.700	2	.008	12.358
228	2	2	.751	6	1.000	3.444	1	.000	27.826
229	1	1	.568	6	.997	4.816	2	.002	16.907
230	1	1	.205	6	.982	8.480	2	.018	16.532
231	2	2	.147	6	.998	9.512	4	.002	21.746
232	1	1	.416	6	1.000	6.067	2	.000	38.310
233	2	2	.923	6	1.000	1.964	1	.000	31.650
234	2	2	.558	6	1.000	4.890	4	.000	34.613
235	2	2	.175	6	.697	8.972	4	.229	11.194
236	7	7	.020	6	1.000	15.028	4	.000	54.724
237	2	2	.783	6	.979	3.202	1	.021	10.872
238	2	2	.879	6	.994	2.404	1	.006	12.758
239	2	2	.969	6	1.000	1.353	1	.000	17.703
240	2	2	.998	6	1.000	.458	1	.000	19.072
241	1	1	.780	6	1.000	3.224	2	.000	26.798
242	2	2	.981	6	.999	1.111	1	.001	15.445
243	1	1	.648	6	1.000	4.209	7	.000	34.271
244	2	2	.847	6	.997	2.691	1	.003	14.296
245	2	2	.853	6	.999	2.633	1	.001	17.308
246	3	3	.937	6	1.000	1.800	2	.000	155.001
247	5	5	.778	6	1.000	3.242	1	.000	42.098
248	7	7	.906	6	1.000	2.143	1	.000	21.651
249	1	1	.572	6	.997	4.785	7	.003	16.125
250	7	7	.301	6	.989	7.220	1	.011	16.149

251	7	7	.033	6	.938	13.691	1	.062	19.113
252	3	3	.409	6	1.000	6.129	1	.000	131.544
253	6	6	.880	6	1.000	2.394	4	.000	135.066
254	5	5	.336	6	1.000	6.839	1	.000	77.243
255	5	5	.884	6	1.000	2.357	1	.000	71.373
256	5	5	.884	6	1.000	2.357	1	.000	71.373
257	5	5	.884	6	1.000	2.357	1	.000	71.373
258	5	5	.642	6	1.000	4.257	1	.000	75.425
259	5	5	.791	6	1.000	3.137	1	.000	73.432
260	5	5	.884	6	1.000	2.357	1	.000	71.373
261	1	7**	.137	6	.720	9.712	1	.280	11.606
262	5	5	.961	6	1.000	1.484	1	.000	55.915
263	5	5	.623	6	1.000	4.398	1	.000	37.656
264	3	3	.882	6	1.000	2.373	1	.000	160.918
265	2	2	.879	6	1.000	2.403	1	.000	32.209
266	2	2	.193	6	.812	8.670	4	.179	11.692
267	7	7	.934	6	1.000	1.839	1	.000	35.609
268	7	7	.406	6	1.000	6.152	1	.000	53.357
269	7	7	.990	6	1.000	.880	1	.000	36.789
270	1	1	.684	6	.999	3.946	2	.001	19.110
271	7	7	.961	6	1.000	1.473	1	.000	37.118
272	5	5	.870	6	1.000	2.489	1	.000	37.435
273	2	2	.861	6	1.000	2.570	1	.000	28.792
274	7	7	.419	6	1.000	6.038	2	.000	23.297
275	2	2	.950	6	1.000	1.641	1	.000	25.455
276	1	1	.940	6	1.000	1.769	2	.000	20.273
277	2	2	.579	6	1.000	4.728	1	.000	29.609

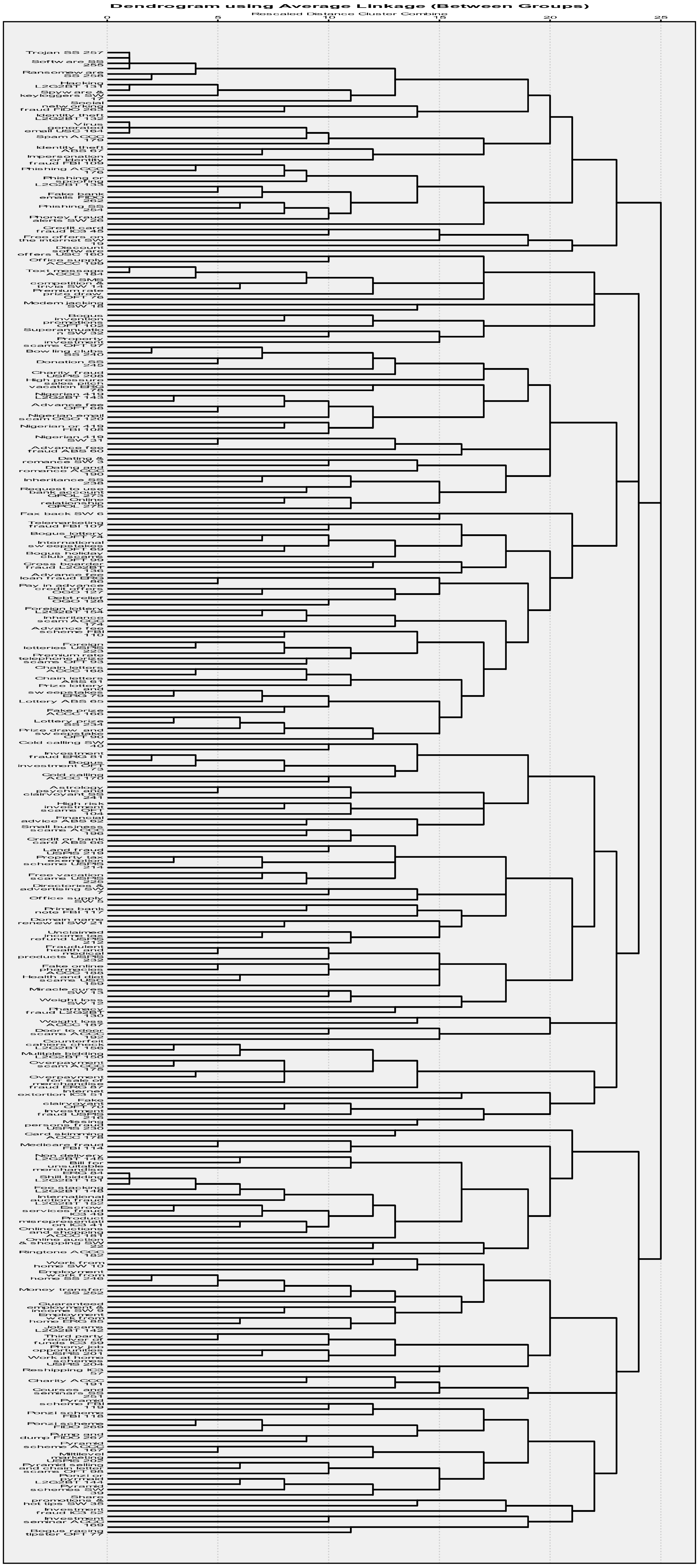


Figure 22: Dendrogram Furthest Neighbour Jaccard Coefficient HCA Model[4]

[4] Dendrograms can be seen in more detail at http://www.icsl.com.au/capability/identity-theft/scams

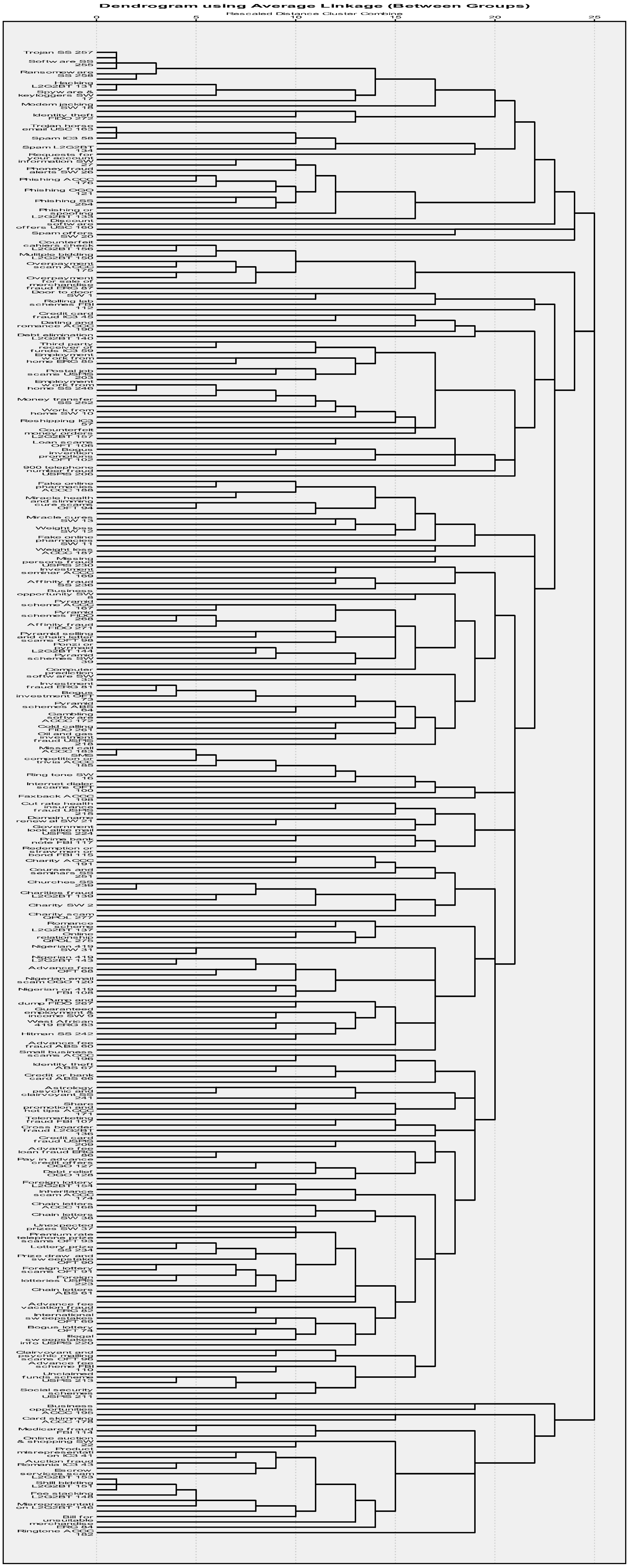


Figure 23: Dendrogram Between Groups Linkage Jaccard Coefficient HCA Model

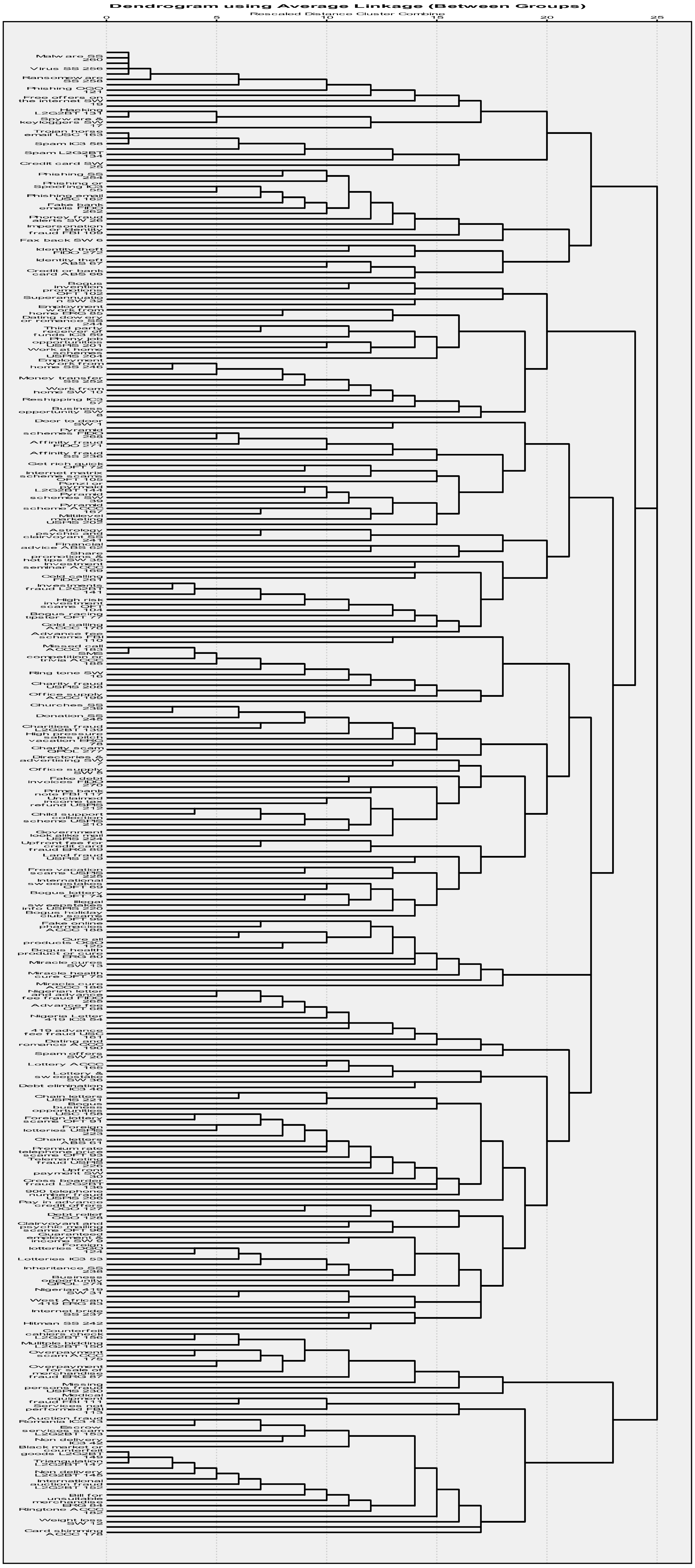


Figure 24: Dendrogram Within Groups Linkage Jaccard Coefficient HCA Model

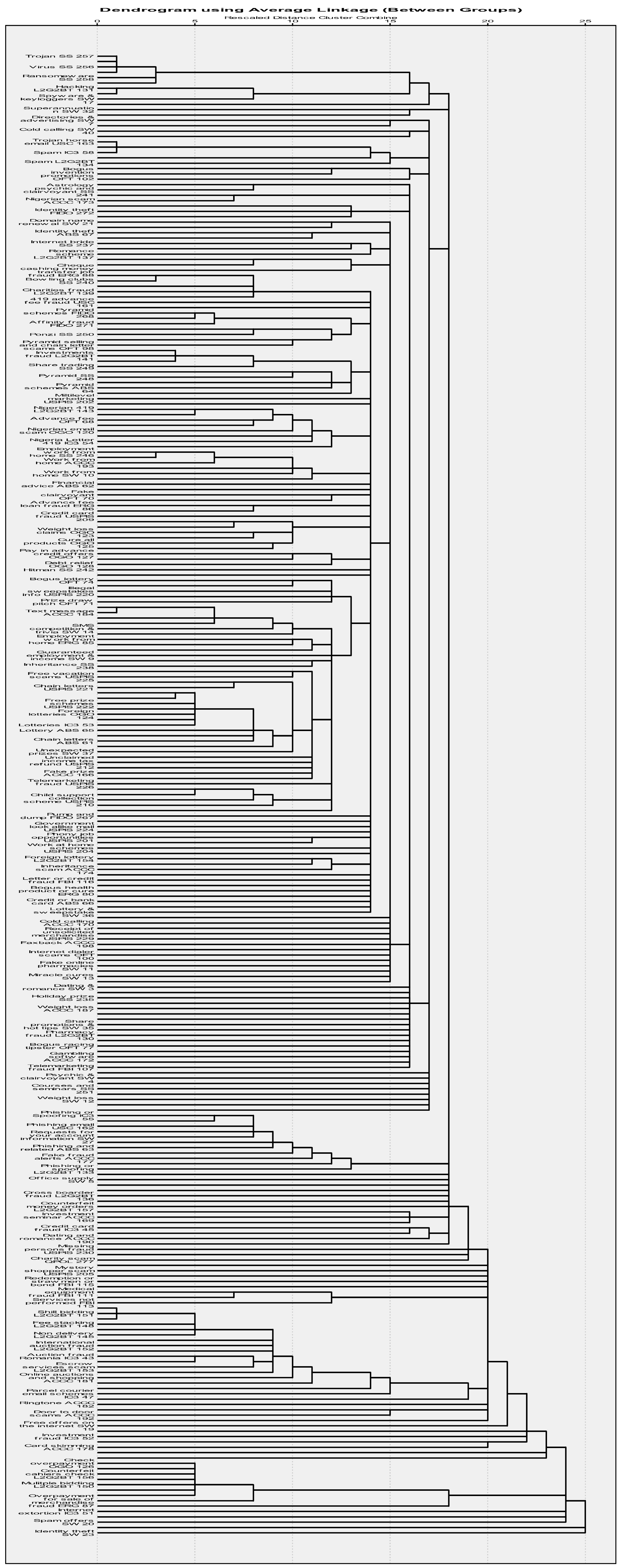


Figure 25: Dendrogram Nearest Neighbour Jaccard Coefficient HCA Model

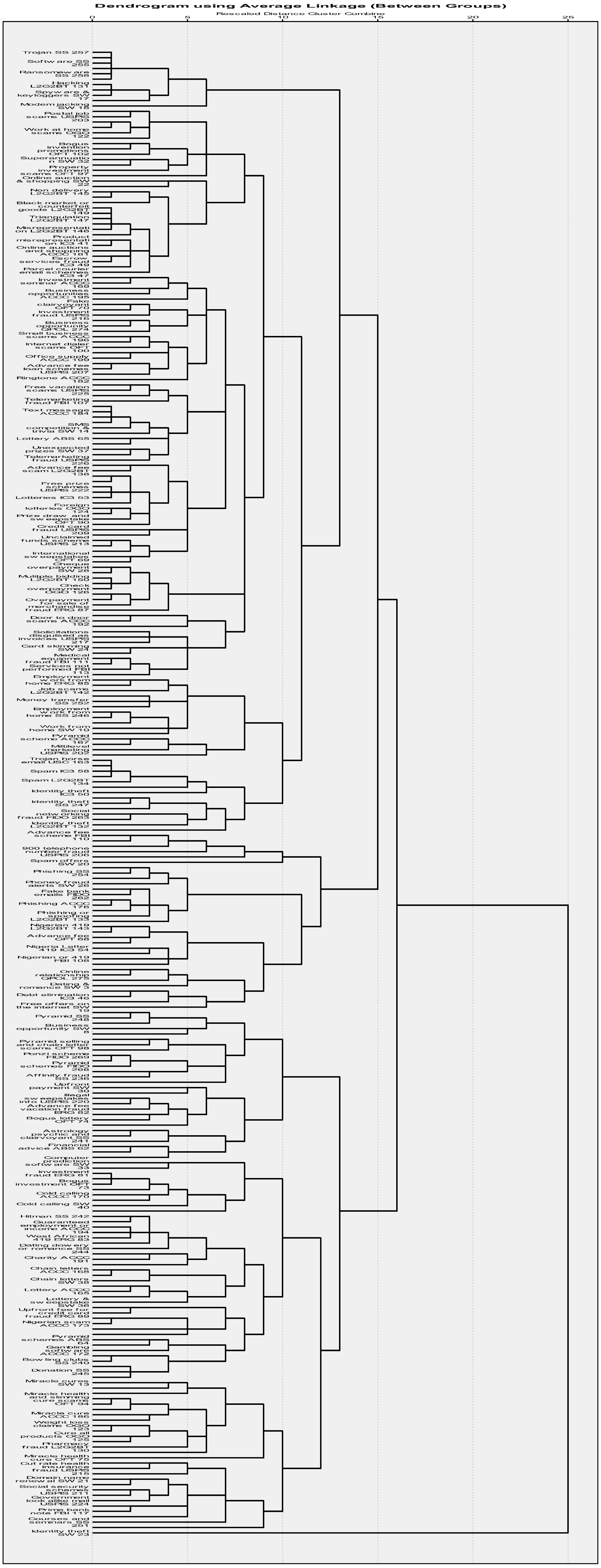


Figure 26: Dendrogram Furthest Neighbour Simple Matching Coefficient HCA Model

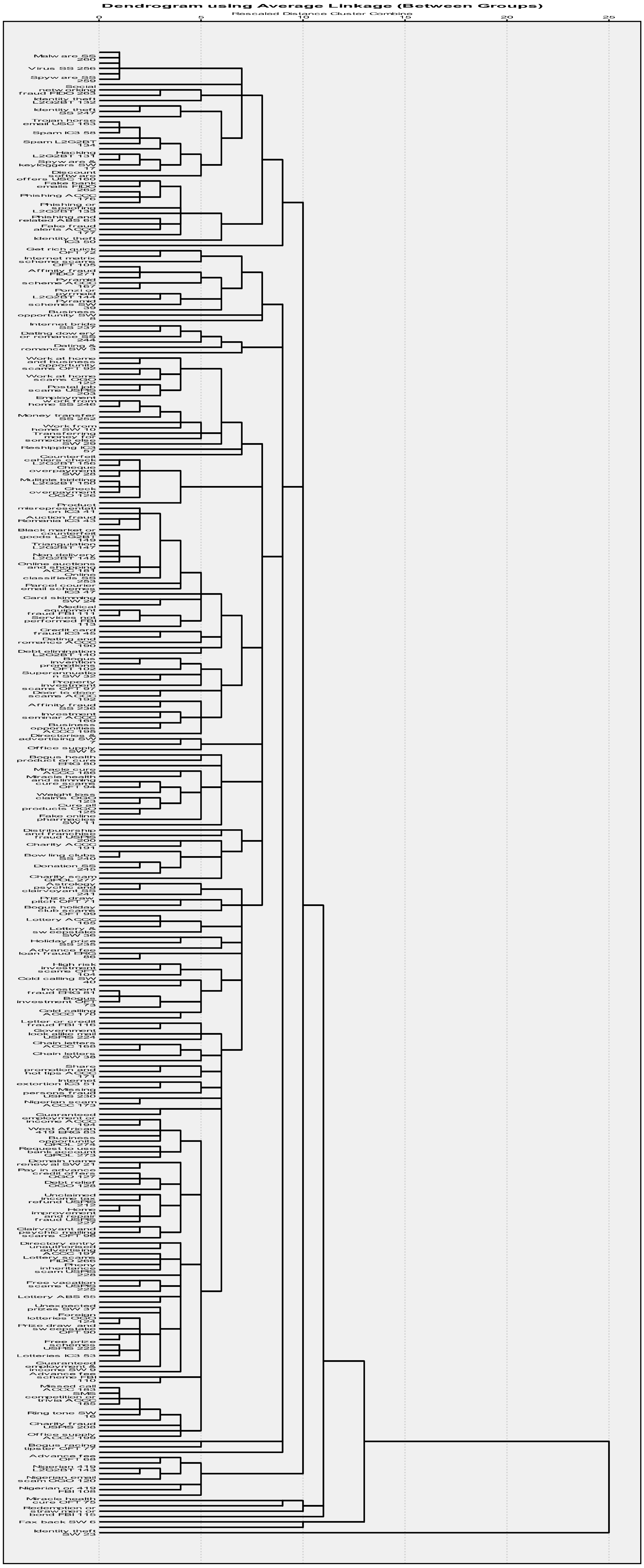


Figure 27: Dendrogram Between Groups Linkage Simple Matching Coefficient HCA Model

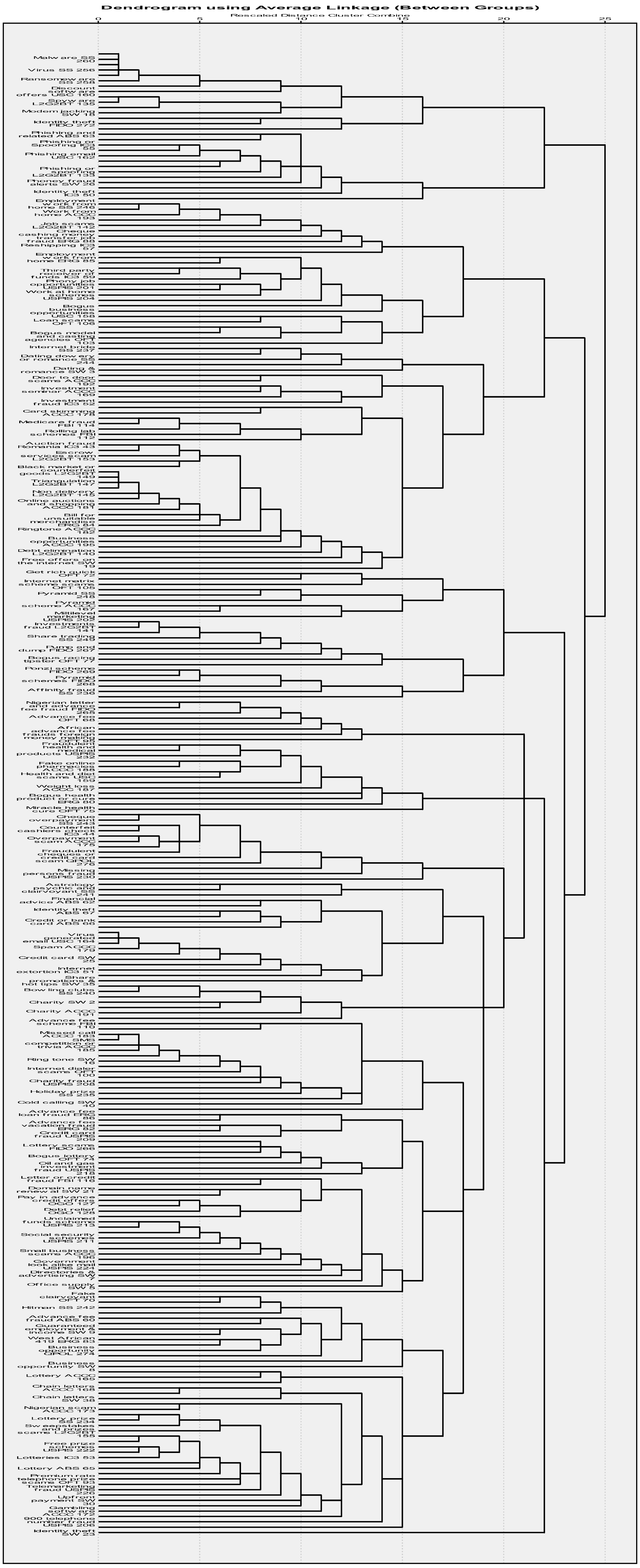


Figure 28: Dendrogram Within Groups Linkage Simple Matching Coefficient HCA Model

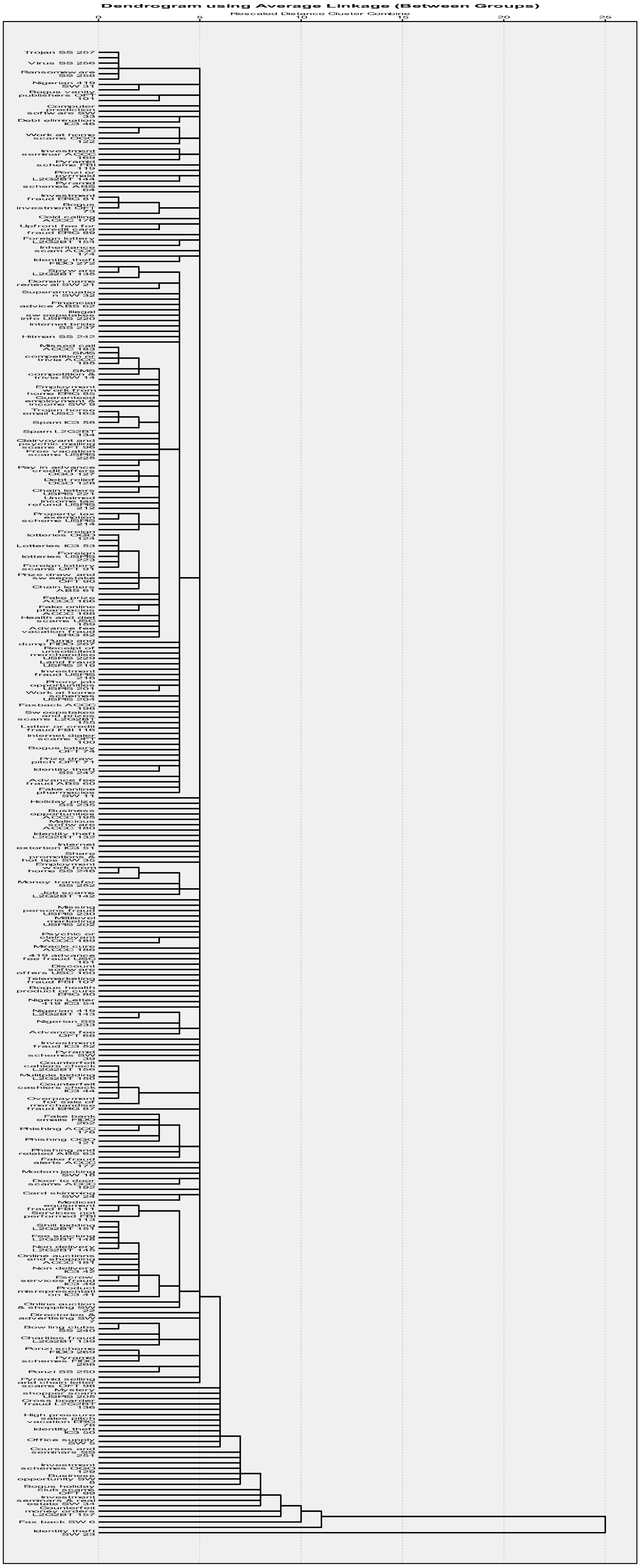


Figure 29: Dendrogram Nearest Neighbour Simple Matching Coefficient HCA Model